



JOURNAL Submitted (16-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Jul 12, 2001 this sequence version replaced gi:12654174.

COMMENT Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalon@bcm.tmc.edu](mailto:villalon@bcm.tmc.edu),  
Villalon, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 3 Row: k Column: 21  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4886442.

FEATURES  
source Location/Qualifiers

1. 1433  
/organism="Homo sapiens"  
/db\_xref="locusid:5861"  
/db\_xref="taxon:9606"  
/clone="MGC:5233 IMAGE:2900705"  
/tissue="Placenta, choriolecarcinoma"  
/clone\_lib="NIH\_MGC\_10"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
179. 796  
/codon\_start=1  
/product="RAB1, member RAS oncogene family"  
/protein\_id="AAH0905.1"  
/db\_xref="gi:12654175"  
/translation="MSEMPEDYDLFKLLIGDSGVKSCLLRFADDTYESTISI  
KQDFKRTILDERTKIKLOIWDPAQOREPTTISVYRGHAGLIIYVDYDQSFNNV  
KQMLQEDIRASENVKLLVGNKCDLTKKAVYDPTAKERADSLGIDFLETSKMNATN  
VEQSEFMMAEIKRMGPATAGGAERSNVKIQSTPVKQSGGGCC"

CDS  
BASE COUNT 432 a 254 c 326 g 421 t

ORIGIN  
Query Match 86.8% Score 1219: DB 9: Length 1433;  
Best Local Similarity 93.2% Pred. No. 2.8e-187;  
Matches 1336; Conservative 0; Mismatches 0; Indels 97; Gaps 2;

QY 3 GCGATAGCTGAGT-GCGGCGGCTGCTGATTTGTCTAGGGAGCGAGTAGGGAAGACG 61  
Db 1 GCGATAGCTGAGTGGCGGCGGCTGCTGATTTGTCTAGGGAGCGAGTAGGGAAGACG 60  
QY 62 TTTGCTCTCCCGGAACACCTATCTCATTCCTTTCTTCGATTACCCGTGCGGAGAG 121  
Db 61 TTTCTCTCTCCCGGAACACCTATCTCATTCCTTTCTTCGATTACCCGTGCGGAGAG 120  
QY 122 TCAGGGGGGGGGTGGGGCAGCAAGGGGGGGGGGGGGGGGGGGGGGGAGCTCAGTGCAT 181  
Db 121 TCAGGGGGGGGGTGGGGCAGCAAGGGGGGGGGGGGGGGGGGGGGGGAGCTCAGTGCAT 180  
QY 182 GTCCAGCATGAATCCGNAATGATTTATTTATTTCAAGTTACTTTCGATTGGCAGTCA 241  
Db 181 GTCCAGCATGAATCCGNAATGATTTATTTATTTCAAGTTACTTTCGATTGGCAGTCA 240  
QY 242 GGTGGAAGTCTTGCTTCTTTAGTTTGAAGTGAATGATATACATACAGAAAGCTACAT 301  
Db 241 GGTGGAAGTCTTGCTTCTTTAGTTTGAAGTGAATGATATACATACAGAAAGCTACAT 300  
QY 302 CAGCAATTTGCTGATTTCAAAATTAAGACATATAGAGTTAAGCGGAAACAAATCA 361  
Db 301 CAGCAATTTGCTGATTTCAAAATTAAGACATATAGAGTTAAGCGGAAACAAATCA 360

Db 301 CAGCAATTTGCTGATTTCAAAATTAAGACATATAGAGTTAAGCGGAAACAAATCA 360  
QY 362 GCTTCAATA----- 371  
Db 361 GCTTCAATATATAGGACACAGCAGCGCAGGAAGATTTCGAACATCACTCCATTTATTA 420  
QY 372 -----GAGTCTTCAATTA 385  
Db 421 CAGAGAGCCCATGATCATATAGTTGTATGATGTGACAGATCAGAGTCCCTCAATTA 480  
QY 386 TGTTAACAGTGGCTGACAGAAATAGATCCGTATCCAGTCAAAATGTCAAATGTT 445  
Db 481 TGTTAACAGTGGCTGACAGAAATAGATCCGTATGCGAGTGAATATGTCAAATGTT 540  
QY 446 GGTAGGGAACAAATGTATCTGACACAAAGAGTGTAGATACACAAAGCAGGAAGA 505  
Db 541 GGTAGGGAACAAATGTATCTGACACAAAGAGTGTAGATACACAAAGCAGGAAGA 600  
QY 506 ATTTGCTGATTCCTTGGAAATTCGTTTGGAAACAGTGTAAAGATGCAACGAATGT 565  
Db 601 ATTTGCTGATTCCTTGGAAATTCGTTTGGAAACAGTGTAAAGATGCAACGAATGT 660  
QY 566 AGAAGAGCTTTCTAGAGATGGCAGCTGAGATTAAGAGGATGGGTCCCGGAGAAC 625  
Db 661 AGAAGAGCTTTCTAGAGATGGCAGCTGAGATTAAGAGGATGGGTCCCGGAGAAC 720  
QY 626 AGCTGTGCTGCTGAGAAGTCCAAATGTTAAATTCAGAGCAGCTCAGTCAAGCAGT 685  
Db 721 AGCTGTGCTGCTGAGAAGTCCAAATGTTAAATTCAGAGCAGCTCAGTCAAGCAGT 780  
QY 686 TGAAGTGTCTGCTAAATTTGCTCCATCTTTTCTCAGACAGATGAATTTGCAATCTG 745  
Db 781 TGAAGTGTCTGCTAAATTTGCTCCATCTTTTCTCAGACAGATGAATTTGCAATCTG 840  
QY 746 AACCAGTGAAGAAACAAATTCGCAATGTCATGATGATGTCGACATCAACAGA 805  
Db 841 AACCAGTGAAGAAACAAATTCGCAATGTCATGATGATGTCGACATCAACAGA 900  
QY 806 TTCTTACCGTCCCAAGAGTCAAGATGTAATGTCATGATGTCATGATGTCATGAT 865  
Db 901 TTCTTACCGTCCCAAGAGTCAAGATGTAATGTCATGATGTCATGATGTCATGAT 960  
QY 866 TCCCTTGACACAGACAGTACTTCAATTTTCAAGACGTTTAAACCTTTGTGTCTGG 925  
Db 961 TCCCTTGACACAGACAGTACTTCAATTTTCAAGACGTTTAAACCTTTGTGTCTGG 1020  
QY 926 TTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 985  
Db 1021 TTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
QY 986 TGGTTAGATATATTTGTTTGTATGATGATGATGATGATGATGATGATGATGAT 1045  
Db 1081 TGGTTAGATATATTTGTTTGTATGATGATGATGATGATGATGATGATGATGAT 1140  
QY 1046 TTTGGAAGATTAATTCAGCATTTTGTATTAACACAGACAGAGTGTGTACATTC 1105  
Db 1141 TTTGGAAGATTAATTCAGCATTTTGTATTAACACAGACAGAGTGTGTACATTC 1200  
QY 1106 CATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1165  
Db 1201 CATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260  
QY 1166 GAGTTATTAATGAAAGATTAACATATCTGATTAATGATGATGATGATGATGAT 1225  
Db 1261 GAGTTATTAATGAAAGATTAACATATCTGATTAATGATGATGATGATGATGAT 1320  
QY 1226 TTTGTGCTGACAGATATTTGATTTGTGACACCTATGTAACAAACCACTGAAGTAT 1285  
Db 1321 TTTGTGCTGACAGATATTTGATTTGTGACACCTATGTAACAAACCACTGAAGTAT 1380  
QY 1286 GTTAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338  
Db 1381 GTTAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1433

[illegible]

D	b	134	AGTCAGGGCGGCGCTGCAGCAAGGCGGCGGTGCAGCGCGCAGCTGCAGTGAC	193
O	y	180	ATGTCACACATAAATCCGAATATGATTAATTATTTAACTCAAGTACTGTATGGGAGATCA	239
D	b	194	ATGTCACACATAAATCCGAATATGATTAATTATTTAACTCAAGTACTGTATGGGAGATCA	253
O	y	240	GGGGTGGAAAAGCTTCGCTCTCTCTTAGTGTTCAGATGTATCATATTCAGAAAAGTCAC	299
D	b	254	GGGGTTGAAAGCTTCGCTCTCTCTTAGTGTTC-----	288
O	y	300	ATCACACAAATGGTGGATTTCAAATTAAGACTATAGAATTAGACGGGAAAAACATC	359
D	b	289	-----	288
O	y	360	AAGCTTCAATAGAGTCTTCAATATGTTTAAAAAGTGGCTGCAGAAATAGATCGTTAT	419
D	b	289	-----AAGTCCCTCAATTAAGTTAAACAGTGGCTGCAGAAATAGATCGTTAT	337
O	y	420	GCCAGTGAATATGTCAACAATTTGTTGTGATAGGAAACAATGTANTGTGACCACAAGAAA	479
D	b	338	GCCAGTGAATATGTCAACAATTTGTTGTGATAGGAAACAATGTANTGTGACCACAAGAAA	397
O	y	480	GTATGAGACTACACAACAGCAGAAGATTTGCTGATTCCTGTGAATTCCTTTTGGAA	539
D	b	398	GTATGAGACTACACAACAGCAGAAGATTTGCTGATTCCTGTGAATTCCTTTTGGAA	457
O	y	540	ACCACTGCTAAGAAATCAACAGATGTGAACAGTCTTTCATAGACGTGCGACGTGAAT	599
D	b	458	ACCACTGCTAAGAAATCAACAGATGTGAACAGTCTTTCATAGACGTGCGACGTGAAT	517
O	y	600	AAAAAGCAATGGGTCCCGAGAACAGCGTGGGTGTGAGAAAGTCCAATGTTAAAT	659
D	b	518	AAAAAGCAATGGGTCCCGAGAACAGCGTGGGTGTGAGAAAGTCCAATGTTAAAT	577
O	y	660	CAGACACCTCAGTCAAGATGACAGGTGGAGGTGTCTGTCAAAATTTGCTTCATCCCTT	719
D	b	578	CAGACACCTCAGTCAAGATGACAGGTGGAGGTGTCTGTCAAAATTTGCTTCATCCCTT	637
O	y	720	TCTCAGACATAATTTGCAATCTGAATGCCAAGTGAAGAAAACAAATTTGCTTGAAATGT	779
D	b	638	TCTCAGACATAATTTGCAATCTGAATGCCAAGTGAAGAAAACAAATTTGCTTGAAATGT	697
O	y	780	ACTGTATGTAGTGCCTACACAAGATCTTACCGCTCCACAAGAGTCAGAGATTTGTAA	839
D	b	698	ACTGTATGTAGTGCCTACACAAGATCTTACCGCTCCACAAGAGTCAGAGATTTGTAA	757
O	y	840	ATGTCATATCTACTACTTTTTTTTTATTCCTTGAGTCAACAAGCAGTACTCATTTTCAG	899
D	b	758	ATGTCATATCTACTACTTTTTTTTTATTCCTTGAGTCAACAAGCAGTACTCATTTTCAG	817
O	y	900	AACCTTTTAAACCTTTGTGTGCGGTGTTTAAATTAATGTGTATTCCTGTGTGTTT	959
D	b	818	AACCTTTTAAACCTTTGTGTGCGGTGTTTAAATTAATGTGTATTCCTGTGTGTTT	877
O	y	960	CCTGATACCAAGACTGTTTCCGCTGGTGGTGAATTAATTTGTTTGAATGTTTAAAT	1019
D	b	878	CCTGATACCAAGACTGTTTCCGCTGGTGGTGAATTAATTTGTTTGAATGTTTAAAT	937
O	y	1020	GGCAATTTAGATGACAGTTTATCTGTATCTGTGAAGATGAATGACCAATTTGTATCAAA	1079
D	b	938	GGCAATTTAGATGACAGTTTATCTGTATCTGTGAAGATGAATGACCAATTTGTATCAAA	997
O	y	1080	CAGCAACAAGAGTGTCTGCACTTTCATGCGATTAAGTTTAGTGAATGTTATATGTAA	1138
D	b	998	CAGCAACAAGAGTGTCTGCACTTTCATGCGATTAAGTTTAGTGAATGTTATATGTAA	1057
O	y	1140	ATCTGATTTCTAGTTCCTCTGTGAGATTTAAATGGAAGATTTACATCTATCTGATTA	1199
D	b	1058	ATCTGATTTCTAGTTCCTCTGTGAGATTTAAATGGAAGATTTACATCTATCTGATTA	1117
O	y	1200	ATAGTTCTTCACTACTGCAATATATTTGTGGCTGCAGAAATTTGTATTTGTTGCACA	1259

DB	1118	ATAGTTTCTCTACTACTCGCANATATATTTGGCGCGGAGAAATATGTAAATTTGTTGCACA	1177
QY	1260	CTATGTAAACAAACACAGTCAGACATATGTTTAATAATAATATGTACTTATTTGGAAGTAAAAA	1319
DB	1178	CTATGTAAACAAACACAGTCAGATATGTTTAATAATAATATGTACTTATTTGGAAGTAAATAT	1237
QY	1320	AAAA	1323
DB	1238	CAAA	1241
RESULT 3			
BC002077			
LOCUS	BC002077	1444 bp	mRNA linear
DEFINITION	Mus musculus, RAB1, member RAS oncogene family, clone MGC:6226		
ACCESSION	IMAGE:3592802, mRNA, complete cds.		
VERSION	BC002077		
KEYWORDS	BC002077.1 GI:12805232		
SOURCE	MGC.		
ORGANISM	house mouse.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 1444)		
JOURNAL	Strausberg, R.		
COMMENT	Direct Submission		
REMARK	Submitted (31-JAN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
COMMENT	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>		
COMMENT	Contact: MGC help desk		
COMMENT	Email: <a href="mailto:cgabs-remail.nih.gov">cgabs-remail.nih.gov</a>		
COMMENT	Tissue Procurement: Gilbert Smith, Ph.D.		
COMMENT	cDNA Library Preparation: Life Technologies, Inc.		
COMMENT	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
COMMENT	DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305		
COMMENT	Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a>		
COMMENT	Contact: (Dickson, Mark) <a href="mailto:mcddpaxil@stanford.edu">mcddpaxil@stanford.edu</a>		
COMMENT	Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a>		
FEATURES	Series: IRAX Plate: 11 Row: a Column: 2		
FEATURES	This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6679586.		
SOURCE	Location/Qualifiers		
SOURCE	1. 1444		
SOURCE	/organism="Mus musculus"		
SOURCE	/db_xref="taxon:10090"		
SOURCE	/map="FVB/N"		
SOURCE	/clone="MGC:6226 IMAGE:3592802"		
SOURCE	/tissue_type="Mammary tumor. Metallochonien-TGF alpha model. 10 month old Virgin mouse. Taken by biopsy."		
SOURCE	/clone_id="NCI CGAP_Mam1"		
SOURCE	/lab_host="DH10B"		
SOURCE	/note="Vector: pCMV-SPORT6"		
SOURCE	196. 813		
SOURCE	/codon_start=1		
SOURCE	/product="RAB1, member RAS oncogene family"		
SOURCE	/protein_id="AAH02077.1"		
SOURCE	/db_xref="GI:12805233"		
SOURCE	/db_xref="locusid:5861"		
SOURCE	/translation="MSSMNPEDYDLFKLLIGDSGVKSCLLFRADDYTESITITITGVDRKIRTELDGNTIKLQIMWDAGQSEFRITSSYKGAAGIIVYDVDDSEFNANNKQMLQIEDIRYASENVNKLQGNKCDLITTKKVVDTAKFADSDIGIFLETSANANNVVEOSFMWAAEIKRKMPGATAGGAESKSNVKGISTPVKSGSGGCC"		
CDS			
BASE COUNT	412 a	270 c	342 g 420 t
Query Match	74.9%	Score 1052.6;	DB 10; Length 1444;

Best Local Similarity		88.3%	Pred. No. 1,9e-160:	
Matches 1256;		Conservative	0;	Mismatches 59; Indels 108; Gaps 6;
QY	16	GGCGGGGCTGCTGATGTGTCTCTAGGGGACGGAGTAGGGGAAAGACGTTTGCTCTCCGGA	75	
Db	27	GGCGGGGCTGCTGATGTGTCTCTAGGGGACGGAGTAGGGGAAAGACGTTTGCTCTCCGGA	85	
QY	76	ACAGCCTATCTCATCTCTCTCTCTGATTTACCCGTGGCGCGGAGTAGAGGGCGCGGCT	135	
Db	86	ACAGCCTATCTCATCTCTCTCTCTGATTTACCCGTGGCGCGGAGTAGAGGGCGCGGCT	145	
QY	136	GGCGGAGCAAGGCG-----GGCGGTGGCGGGGGCGGAGCGAGTGCATGATGTCCACGA	189	
Db	146	GGGTAGCAAGGCGCGGTGTGGCGGGCGGCGGCGGAGCTGCGAGTGCATGATGTCCACGA	205	
QY	190	TGAATCCCGAATATGATTATTTATTTCAAGTTACTTCTGATTGGCGAGTCAAGGGTTGGA	249	
Db	206	TGAATCCCGAATATGATTATTTATTTCAAGTTACTTCTGATTGGCGAGTCTGGGGTTGGA	265	
QY	250	AGCTTGCCCTTCTTTAGGTTTGCAAGTATATACATATACAGAAAGCTATCATCACACA	309	
Db	266	AGCTTGCCCTTCTCTTGTGGTTTGCAAGTATATACATATACAGAAAGCTATCATCACACA	325	
QY	310	TTGGTGGGATTTCAAAATTAAGACTATAGATTAGAGGGGAAACATTCAGCTTCAA	369	
Db	326	TTGGTGGGATTTCAAGATACGACTATAGATTAGATGGGAAACATTCAGCTTCAA	385	
QY	370	TA-----	371	
Db	386	TATGGGACACAGCAGGCGGAGAAATTTGCAGCAATCACTTCAGTTATTACAGAGAG	445	
QY	372	-----GAGTCTTCAATATGTTAAAC	393	
Db	446	CCCATGGCATCATAGTTGTATGTATGTATGTACAGATCAGAGTCTTCAATATGTTAAAC	505	
QY	394	AGTGCTCGAGAAATAGATGCTATGTCAGAGAAATGTCAACAAATGTTGGTAGGA	453	
Db	506	AGTGCTCGAGAGATAGTCTCTACGCCAGTGAAGAAATGTCAACAAATGTTGGTAGGA	565	
QY	454	ACAAATGTGATCTGACCAACAAAGAAAGTAGTAGACTACACACAGCGAAGAAATTTGCTG	513	
Db	566	ACAAATGTGACTGACCAACAAAGAAAGTAGTAGACTACACACAGCGAAGAAATTTGCGAG	625	
QY	514	ATTCCCTTGGAAATTCCTTTTGGAAACAGTGTCAAGTATGCAACGAATGTAGAACAGT	573	
Db	626	ATTCCCTTGGAAATTCCTTTTGGAAACAGTGTCAAGTATGCAACGAATGTAGAACAGT	685	
QY	574	CTTTCATGCGATGGCAGCGTGTGAATTAAGAACGAATGGGTCCGAGACAACAGCTGGTG	633	
Db	686	CTTTCATGCGATGGCAGCGTGTGAATTAAGAACGAATGGGTCTTGAGCTACAGCTGGTG	745	
QY	634	GTCGTGAGAAAGTCCATGTTTAAATTTCAAGACACTCCAGTCAAGCAGTCAAGTGGAGTT	693	
Db	746	GTCGTGAGAAAGTCCATGTTTAAATTTCAAGACACTCCAGTCAAGCAGTCAAGTGGAGTT	805	
QY	694	GCTGCTAAATTTGCTCTCATCTTTTCTCACAGCAATGATTTGCAATCTGAAACCAAG	753	
Db	806	GCTGCTAAATTTGCTCTCATCTTTTCTCACAGCAATGATTTGCAATCTGAAACCAAG	865	
QY	754	TCAAAAAACAAATTTGCTCGAATTTGATCTGTATGTAGTGGCACTCAACAGATTTCTTACC	813	
Db	866	TCAAAAAACAAATTTGCTCGAATTTGATCTGTATGTAGTGGCACTCAACAGATTTCTTACC	925	
QY	814	GTCGTACAAAGGTGAGAGATGTGAATGGTCAATACTGAC-TTTTTTTTTTATTCCTTG	872	
Db	926	GTTTCCAC- AAGGTAGAGATTTGAATGGTCAATACTGACATTTTTTTTTTATTCCTTG	984	
QY	873	ACTCAAGACAGCTAATCTATTTTCAGAACTGTTTAAACCTTTGTGCTGGTTATAA	932	
Db	985	ACTCAAGACAGCTAATCTATTTTCAGAACTGTTTAAACCTTTGTGCTGGTTATAA	1044	
QY	933	AATTAATGTGTAACTCTTGTGCTTCTCTGATACAGACTGTTTCCGCTGGTTAG	992	



Df	1045	AATAAAGCTGTAAVCCCTTGGCTTCCTGATACCAAGATCCTTTCCCGTGTTGGTTAG	1104
Oy	993	AATAATTTTTGTTTGAATGATTATATATGATGATGTTAGATGCAGCTTATGCTCTGAAA	1092
Df	1105	AATAATTTTTGTTTGAATGATTATATGATGATGTTAGATGCAGCTTATGCTCTGAAA	1164
Oy	1053	GATGAAGTTCCGCCATTTTGTATCAAACAAGCACAGCATGCTGTCTCATCTCCAGCAT	1112
Df	1165	GATGAAGTTCCGCCATTTTGTATCAAACAAGCACAGCATGCTGTCTCATCTCCAGCAT	1224
Oy	1113	AAAGTTTAGCAGATGTTATATGTAAAGATCTGATTTGCTAGTCTTCTCTTGAGACTAT	1172
Df	1225	AAAGTTTAGCAGATGTTATATGTAAAGATGATGCTAGTCTTCTCTTGAGACTAT	1284
Oy	1173	AAATGGAAGATTACACATCTGATTTAATAGTCTTCTCATCTGCATTAATATTTGTGG	1232
Df	1285	AAATGGAAGATTACACATCTGATTTAATAGTCTTCTCATCTGCATTAATATTTGTGG	1344
Oy	1233	CTCCAGAATATTTGATTTGTTGACACTATGTAACAAACAATGGAAGATATGTTAAT	1292
Df	1345	CTCCAGAATATTTGATTTGTTGACACTATGTAACAA---AACGGAAGATATGTTAAT	1401
Oy	1293	AAATATGTACTATTTGGAAGTAAAAAATAAAAAAAAA	1335
Df	1402	AAATATGTACTATTTGGAAGTAAAAAATAAAAAAAAA	1444
RESULT 4			
LOCUS	HSMB00023		
DEFINITION	Homo sapiens mRNA; cdna DKFP564B163 (from clone DKFP564B163); complete cds.		
ACCESSION	AF050268		
VERSION	AL050268.1 GI:4886442		
KEYWORDS	human;		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Wiemann,S., Well,B., Wellenreuther,R., Gassenhuber,J., Glassl,S., Ansoerje,W., Boecker,M., Bloeker,H., Baursachs,S., Blum,H., Lader,J., Duesterhoeft,A., Beyer,A., Koehler,K., Strack,N., Meves,H.W., Ottenwaeldeit,B., Obermaier,B., Tampe,J., Heubner,D., Wandt,U.R., Korn,B., Klein,M. and Poustka,A.		
TITLE	Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs		
JOURNAL	Genome Res. 11 (3), 422-435 (2001)		
PUBMED	11230166		
REFERENCE	2 (bases 1 to 1208)		
AUTHORS	Wandut,R., Heubner,D., Meves,H.W., Gassenhuber,J. and Wiemann,S.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-MAR-1999) MIPS, Am Klopfersplitz 18a, D-82152 Martinsried, GERMANY		
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ). Email s.wiemann@dkfz-heidelberg.de; Consortium of the German Genome Project. This clone (DKFP564B163) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcentzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY. Email: clonerzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cdna/. LocalID/Qualifiers 1..1208		
SOURCE	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFP564B163" /tissue_type="brain" /clone_lib="564 (synonym: hfbz2). Vector pAMP1; host Kl-2blue; sites NotI + SalI" /dev_stage="fetal" 191..580		
gene			



QY 797 TACAGACATCTTACGCTCCACAAAGGTCAGAGATTTGTAATGGTCATACGAC-T 855  
|||||  
Db 843 TACACACATCTTACGCTTCCAC-AGAGTCAGACATTTGTAATGGTCATACGAC-T 901  
|||||  
QY 856 TTTTCTTATCCCTGACCTCAAGACACTTACCTGATTTTCAAGACTGTTTTAACTT 915  
|||||  
Db 902 TTTTCTTATCCCTGACCTCAAGACCTTACCTGATTTTCAAGACTGTTTTAACTT 961  
|||||  
QY 916 TGTGTGCTGTTTATAAATATGTTGTAATCCCTGTTCTCCATACCACTGT 975  
|||||  
Db 962 TGTGTGCTGTTTATAAATATGTTGTAATCCCTGTTCTCCATACCACTGT 1021  
|||||  
QY 976 TTTCCGCTGCTGTTGTAATATTTGTTGTAATTTTATTTGCAATGTTTAACTGTC 1035  
|||||  
Db 1022 TTTCCGCTGCTGTTGTAATATTTGTTGTAATTTTATTTGCAATGTTTAACTGTC 1081  
|||||  
QY 1036 AGCTTACCTCTTCAAGATGAGTACGCTATTTGTAACAAACAGCAGCACTGTC 1095  
|||||  
Db 1082 GGTCTTACCTCTTCAAGATGAGTACGCTATTTGTAACAAACAGCAGCACTGTC 1141  
|||||  
QY 1096 TGTCTACCTTCCATGATTAAGTTTGTAGATGTTTATGTAGATGTTGCTAGTT 1155  
|||||  
Db 1142 TGTCTACCTTCCATGATTAAGTTTGTAGATGTTTATGTAGATGTTGCTAGTT 1201  
|||||  
QY 1156 CTTCCTTTGTAGATTTTAAATGGAAGATTACCTATCTGATTAATGTTTCTTCACT 1215  
|||||  
Db 1202 CTTCCTTTGTAGATTTTAAATGGAAGATTACCTATCTGATTAATGTTTCTTCACT 1261  
|||||  
QY 1216 CTGCATTAATTTGTGCTGTCAGATATGTAATTTGTGCACTATGTACAAACAA 1275  
|||||  
Db 1262 CTGCATTAATTTGTGCTGTCAGATATGTAATTTGTGCACTATGTACAAACAA 1318  
|||||  
QY 1276 CTGAAGATATGTTTAAATATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1323  
|||||  
Db 1319 CTGAAGATATGTTTAAATATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1366  
|||||

RESULT 6  
AF226873 2686 bp mRNA linear ROD 13-FEB-2000  
LOCUS  
DEFINITION Mus musculus small GTP-binding protein RAB1A mRNA, complete cds.  
ACCESSION AF226873  
VERSION AF226873.1 GI:6969621  
KEYWORDS  
SOURCE  
ORGANISM  
Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Wu, G. and Dorn, G.W. II.  
TITLE Mouse RAB1A, member of RAS oncogene family mRNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2686)  
AUTHORS Wu, G. and Dorn, G.W. II.  
TITLE Direct Submission  
JOURNAL Submitted (19-JAN-2000) Internal Medicine, University of  
Cincinnati, 221 Bethesda Avenue, Cincinnati, OH 45267-0590, USA

FEATURES  
source  
Location/Qualifiers  
1..2686  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
29..646  
/note="member of RAS oncogene family"  
/codon\_start=1  
/product="small GTP-binding protein RAB1A"  
/protein\_id="AF33844.1"  
/db\_xref="GI:6969621"  
/translation="MSSMPEYDYLFRKLLIGSGYKSCULLRPADTYTESIST  
GVDFKRIEDGKTKIDQIMDTAGQEFRTISSTYRAGHITVVDVDDSEFNV  
KQMLQEDRTASEVNLVGNKCDLTKKVVDTTAKFADSLGIFLETSAKATN  
VQSPFMVAIRKRWKPGATVGGERSNVKIQSTPVKQSGGCC"

BASE COUNT 778 a 520 c 563 g 825 t

## ORIGIN

Query Match 68.0%; Score 955; DB 10; Length 2686;  
Best Local Similarity 88.5%; Pred. No. 1,1e-144;  
Matches 1123; Conservative 0; Mismatches 45; Indels 101; Gaps 4;

QY 152 CGGTGGCGGGCGGCGGAGCTGACATGATGTCAGCATGAAATCCGAATATGATATT 211  
|||||  
Db 1 CGCGGGCGGGCGGCGGAGCTGACATGATGTCAGCATGAAATCCGAATATGATATT 60  
|||||  
QY 212 ATTCAAGTTCTTGATATGAGGAGCTGACAGGAGTGAAGTTCGCTCTCTAGATT 271  
|||||  
Db 61 ATTCAAGTTCTTGATATGAGGAGCTGACAGGAGTGAAGTTCGCTCTCTAGATT 120  
|||||  
QY 272 TGCAGATGATATATACAAAGATCAATCATACAGCAATTTGTGATTTCAAAATAG 331  
|||||  
Db 121 TGCAGATGATATATACAAAGATCAATCATACAGCAATTTGTGATTTCAAAATAG 180  
|||||  
QY 332 AACTATAGATTGACGGGAAACAAATCAAGCTTCAATA- 371  
|||||  
Db 181 AACTATAGATTGATGGGAAACAAATCAAGCTTCAAGATATGGGACACAGCGCAGGA 240  
|||||  
QY 372 ----- 371  
|||||  
Db 241 AAGATTTCGAACATCACTTCCAGTTATTCAGAGAGCCCATGCAATATGTTGTA 300  
|||||  
QY 372 ----- 415  
|||||  
Db 301 TGATGTGACGATGACGAGAGTCTTCAATTAACGTTAAACAGTGCCTCAGAGATGATCG 360  
|||||  
QY 416 TTATGCGAGTAAATGTCACAAATTTGGTAGAGAAATATGATCTACACAAA 475  
|||||  
Db 361 CTAGCGCAGTAAATGTCACAAATTTGGTAGAGAAATATGATCTACACAAA 420  
|||||  
QY 476 GAAAGTAGTACCTACACAGAGGAAATTTCTGATTTCCCTTGGAATTCGTTTT 535  
|||||  
Db 421 GAAAGTAGTACCTACACAGAGGAAATTTCTGATTTCCCTTGGAATTCGTTTT 480  
|||||  
QY 536 GAAACCACTGTGTAAGATGCAAGATGTAGACAGTCTTTCATGACANTGGCAGTGA 595  
|||||  
Db 481 GAAACCACTGTGTAAGATGCAAGATGTAGACAGTCTTTCATGACANTGGCAGTGA 540  
|||||  
QY 596 GATTAAAAAGCGAATGGGTCCCGGAGCAACAGCTGTGTGCTGAGAAATCCAAATGTTAA 655  
|||||  
Db 541 GATTAAAAAGCGAATGGGTCCCGGAGCAACAGCTGTGTGCTGAGAAATCCAAATGTTAA 600  
|||||  
QY 656 AATTCAGAGCACTCCAGTCAAGCAAGTACAGTGGAGTTGCTCTAAATTTGCTCATC 715  
|||||  
Db 601 AATTCAGAGCACTCCAGTCAAGCAAGTACAGTGGAGTTGCTCTAAATTTGCTCATC 660  
|||||  
QY 716 CTTTCTCAGACAGTAATTTGCAATCTGAACCCCAAGTGAAGAAACAAATTTGCTGAA 775  
|||||  
Db 661 CTTTCTCAGACAGTAATTTGCAATCTGAACCCCAAGTGAAGAAACAAATTTGCTGAA 720  
|||||  
QY 776 TTTGACGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 835  
|||||  
Db 721 TTTGACGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 779  
|||||  
QY 836 GTAATGCTCAATCTGAC-TTTTTTTTATTCCTTGACTCAAGACAGTCACTCAAT 894  
|||||  
Db 780 GTAATGCTCAATCTGAC-TTTTTTTTATTCCTTGACTCAAGACAGTCACTCAAT 839  
|||||  
QY 895 TTTGGAACGTGTTTAAACCTTTGTGCTGCTGTTTATAAATATGTTAATCTGTT 954  
|||||  
Db 840 TTTGGAACGTGTTTAAACCTTTGTGCTGCTGTTTATAAATATGTTAATCTGTT 899  
|||||  
QY 955 GCTTCTGATACAGACTGTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1014  
|||||  
Db 900 GCTTCTGATACAGACTGTTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959  
|||||  
QY 1015 AATTTGCAATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1074  
|||||  
Db 960 AATTTGCAATGTTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1019  
|||||

OY	1075	TCAAACACACACAGAGGTCTGTGCATCTTCATCAATAAGTTT	AGTGAAGATGTTAAAT	1134
Db	1070	TCACACACACACACAGGTCTGTGTAGTTTCCACGCATTAAGTT	ATAGTACAGCTTAAT	1079
OY	1135	GTAAGATCTGATTTGTCATGTTCTTCCCTTGAGAGTTTAATGA	AGATTACACTATCT	1194
Db	1080	GTAAGATCTGATTTGTCATGTTCTTCCCTGTAGAGTTTAATAA	AGGAAGATTACACTATCT	1139
OY	1195	GATTAAATAGTTTCTTCATACCTCGCATTAATTTGGCTGCAGA	ATATTGAATTTGTT	1254
Db	1140	GATTAAATAGTTTCTTCATACCTCGCATTAATTTGGCTGCAGA	ATATTGAATTTGTT	1199
OY	1255	GCACACTATGTAAACAAACAACCTGGAAGATATGTTTAATAAT	ATTTGTACTTAATTTGGAAGT	1314
Db	1200	GCACACTATGTAAACAAACAACCTGGAAGATATGTTTAATAAT	ATTTGTACTTAATTTGGAAGT	1256
OY	1315	AAAAAAAAA	1323	
Db	1257	AATATCAAA	1265	

RESULT	7
AC123143	
LOCUS	
DEFINITION	AC123143 151988 bp DNA linear HTG 17-JUL-2002 Rattus norvegicus clone CH230-247B24, *** SEQUENCING IN PROGRESS ***, 50 unordered pieces.

ACCESSION	AC123143
VERSION	AC123143..2
KEYWORDS	HTG; HTGS; PHASE1.
SOURCE	Norway ratl.
ORGANISM	Rattus norvegicus

REFERENCE  
1 (bases 1 to 151988)  
Munhy D.M., Adams C., Adlo-Oduola B., Ali-osman F.R., Allen C.,  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

Albarracks, S.L., Amarantunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbato, J., Benton, J., Blimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowle, S., Brievara, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Deederich, D.A., Delaney, K.R., Delgado, O., Dem, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabelli, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrrell, J., H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, H., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivel, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Ilean, C., Liu, J., Liu, W., Lounsged, H., Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Marindale, A., Martinez, E., Massey, E., McWhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miller, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Oguh, M., Okunodu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Qulies, M., Ren, Y., Rives, M., Rojas, A., RojudoKan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamejis, A., Tamejis, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,

TITLE  
 JOURNAL  
 REFERENCE  
 TITLE  
 AUTHORS  
 JOURNAL

Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Direct Submission  
 unpublished  
 2 (bases 1 to 151988)  
 Worley, K.C.  
 Direct Submission  
 Submitted (29-MAY-2002) Human Genome Sequencing Center, Department  
 JOURNAL

REFERENCE  
3 (bases 1 to 151988)  
Baylor Plaza, Houston, TX 77030, USA  
OI Molecular and Human Genetics, Baylor College of Medicine, One

**AUTHORS** WOILEY, A. C.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (17-JUL-2002) Human Genome Sequencing Center, Department

COMMENT  
O'Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 4, 2002 this sequence version replaced gi:212338695.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

-----  
Project Information  
Contact: hgsc-help@bcm.tmc.edu  
Web site: <http://www.hgsc.bcm.tmc.edu/>

```

Center project name: GMOF
Center clone name: CH230-247B24
-----
Summary Statistics

```

Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329

Consensus quality:	108541 bases at least Q40
Consensus quality:	113010 bases at least Q30
Consensus quality:	115635 bases at least Q20

-----  
 \* NOTE: Estimated insert size may differ from sequence length.  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))

\* NOTE: This is a 'working draft' sequence. It currently consists of 50 contigs. The true order of the pieces is not known and their order in this sequence record is

- \* arbitrary. Gaps between the contigs are represented as
- \* runs of N, but the exact sizes of the gaps are unknown.
- \* This record will be updated with the finished sequence.

\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1041: contig of 1041 bp in length

*	1042	1141: gap of unknown length
*	1142	2641: contig of 1500 bp in length
*	1143	2741: gap of unknown length
*	2642	

*	3932	5616: contig of 1685 bp in length
*	3832	
*	3831	3831: contig of 1090 bp in length
*	2742	

*	5617	5716: gap of unknown length
*	5717	6732: contig of 1016 bp in length
*	6733	6832: gap of unknown length

*	6833	8189: contig of 1357 bp in length
*	8289	gap of unknown length
*	8290	9410: contig of 1121 bp in length

*	9411	9510: gap of unknown length
*	9511	10654: contig of 1144 bp in length
*	10655	10754: gap of unknown length

*	10755	11847: contig of 1093 bp in length
*	11848	11947: gap of unknown length
*	11948	13492: contig of 1545 bp in length

*	13592: gap of unknown length
*	13593
*	14687: contig of 1095 bp in length
*	14787: gap of unknown length
*	14688

*	14788	16151: contig of 1364 bp in length
*	16152	16251: gap of unknown length
*	16252	18133: contig of 1882 bp in length

*	18134	gap of unknown length
*	18234	contig of 1632 bp in length
*	19865	gap of unknown length

\* 19966 21209: contig of 1244 bp in length

[illegible]

OY	863	TATTCCTTGACCTCAAGACAGCTACTCTCATTTTCAGAACTGTTTAAACCTTCTGTGC	922
Db	76881	TATTCCTTGACCTCAAGACAGCTACTCTCATTTTCAGAACTGTTTAAACCTTCTGTGC	76940
OY	923	TGCTTTATAAATAATATGCTGTAACTCTTGTGCTTCTCTGATACAGACTGTTCCCGT	982
Db	76941	TGCTTTATAAATAATATGCTGTAACTCTTGTGCTTCTCTGATACAGACTGTTCCCGT	77000
OY	983	GGTTCGTTAGAAATATATTTTGTGTTGATGTTATATTTGGCATGTT-----	1027
Db	77001	GGTTCGTTAGAAATATATTTTGTGTTGATGTTATATTTGGCATGTTAGATGTTGGGTTTA	77060
OY	1028	-----	1027
Db	77061	GTTGGGGATTATAGCTCAGTGTAGAGCCCTTGCCTAGCAAGCGCAAGCCCTGGGTTCCG	77120
OY	1028	-----TAGATGTCAAGGTTTATAGTCTCTCTGAAAGATGAAGTTCAAGC	1066
Db	77121	TCCCCAGCTCGCAAAAAAATATAGATGTGTGGGTTTAGTCTTCTGAAAGATGAAGTTCAAGC	77180
OY	1067	ATTTTGTATCAAAACAGCAACAGCAGTGTCTGTCACTTTCATGATCAATTAAGTTTACTGAGA	1126
Db	77181	ATTTTGTATCAAAATGTGCAACACACAGTGTCTGTCACTTTCATGATCAATTAAGTTTACTGAGA	77240
OY	1127	TGTTTATGTAGTAGATCTATTTGTCAGTGTCTCCCTGTAGAGTATTAATGGAAGATTA	1186
Db	77241	CATTATATGTAGTAGATCTATTTGTCAGTGTCTCCCTGTAGAGATTAATTAATGGAAGATTA	77299
OY	1187	CACATCTCGATTAATATAGTTTCTTCATCTCTGATATATTTTGTGGTCGACATATTTGT	1246
Db	77300	CACATCTCGATTAATATAGTTTCTTCATCTCTGATATATTTTGTGGTCGACATATTTGT	77359
OY	1247	AATTTGTTGACACTATGTAAACAAACAACCTGAGATATGTTTAAATATATTGTACTTA	1306
Db	77360	AATTTGTTGACACTATGTAAACA---AAGTGAAGACATGTTTAAATATATTGTACTTA	77416
OY	1307	TTTGAAGTAAAAAAA 1323	
Db	77417	TCGGAAGTAAATATCAAA 77433	

RESULT 8	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE
AC007318/c	AC007318	Homo sapiens BAC clone RP11-420C9 from 2, complete sequence.	204230 bp	DNA	linear	PRI 07-NOV-2001	
	AC007318	Homo sapiens BAC clone RP11-420C9 from 2, complete sequence.	204230 bp	DNA	linear	PRI 07-NOV-2001	
	AC007318..4	GI:13435282					
	HTG.						
	Homo sapiens.						
	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
	1 (bases 1 to 204230)						
	Sulston, J.E. and Waterston, R.						
	Toward a complete human genome sequence						
	Genome Res. 8 (11), 1097-1108 (1998)						
	99063792						
	PUBMED						
	9847074						
	2 (bases 1 to 204230)						
	Wollam, A., Hawkins, M., Kalicki, J., Harkins, R. and Lehnert, L.						
	The sequence of Homo sapiens BAC clone RP11-420C9						
	unpublished						
	3 (bases 1 to 204230)						
	Waterston, R.H.						
	Direct Submission						
	Submitted (15-APR-1999) Genome Sequencing Center, Washington						
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,						
	MO 63108, USA						
	4 (bases 1 to 204230)						
	Waterston, R.H.						
	Direct Submission						
	Submitted (23-MAR-2001) Genome Sequencing Center, Washington						
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,						

REFERENCE	MO 63108, USA
AUTHORS	5 (bases 1 to 204230)
TITLE	Waterston, R.
JOURNAL	Direct Submission
REFERENCE	Submitted (09-AUG-2001) Department of Genetics, Washington
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
TITLE	6 (bases 1 to 204230)
JOURNAL	Waterston, R.
COMMENT	Direct Submission
	Submitted (07-NOV-2001) Department of Genetics, Washington
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
	On Mar 23, 2001 this sequence version replaced g1:7630964.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:** Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:** The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.-Y., Zhao, B., Frenken, E., Tatenoe, M., Catanesse, J. J. and de Jong, P. J. (1988) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletzer de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the left is RP11-350H15; the clone sequenced to the right is AC034121. Actual start of this clone is at base position 1 of RP11-420C9; actual end is at base position 204230 of RP11-420C9.

The sequence Rp11-420c9 from base position 44230 to 46506 consists of PCR product from clone DNA. Base position 44700 to 44774 represents sequence of a single plasmid subclone derived from PCR amplification of clone DNA.

```

FEATURES
source      Location/Qualifiers
1. .204230

```

```
repeat_region      179.486  
                    /rpt_family="Alu"  
repeat_region      890.1082  
                    /rpt_family="MER1_type"  
repeat_region      1498.1789
```

	misc_feature	6663..6949	/note=*similar to EST T07351 (NID:g138500) *
	misc_feature	6861..7737	/note=*similar to EST A0124391 (NID:g10949107) "
	misc_feature	7006..7454	/note=*similar to EST AA825455 (NID:g2896766) oee4t08.s1"
	misc_feature	7130..7950	/note=*similar to EST BF242402 (NID:g1156329) "
	repeat_region	7171..7407	
	misc_feature	7320..7336	73PC_family "Alu"
	misc_feature	7501..8108	/note=*similar to EST AV703078 (NID:g10719408) *
	misc_feature	7650..7920	/note=*similar to EST BE279740 (NID:g9154735) "
	misc_feature	7790..8251	/note=*similar to EST AV703078 (NID:g10719408) "
	misc_feature	7945..8533	/note=*similar to EST AA25857 (NID:g1891513) zsz7h08.r1"
	misc_feature	8047..8633	/note=*similar to EST AV601747 (NID:g9724069) "
	misc_feature	8065..8647	/note=*similar to EST AME50283 (NID:g7195711) *
	misc_feature		/note=*similar to EST AU149130 (NID:g11010651) "
	Query Match	58.2%; Score 817.2; DB 9; Length 204230;	
	Best Local Similarity	99.6%; Pred. No. 2,1e-122;	
	Matches 819; Conservative	0; Mismatches 3; Indels 0; Gaps 0;	
OY	502	AGCAATTTGCTGAATTCCCTTGAAATTCGGTTTTTGGAAACCGACTGTAGAATGCAMACA	561
Db	10336	AGCAATTTGCTGAATTCCTTGGAATTCCTTTTGGAAACCGACTGTAGAATGCAMACA	10277
OY	562	ATGTAGAAGCTCTTTCATGACAGTGGCAGCTGAGATTTAAAAAGCGAATGGGTCCGGAG	621
Db	10276	ATGTAGAAGCTCTTTCATGACAGTGGCAGCTGAGATTTAAAAAGCGAATGGGTCCGGAG	10217
OY	622	CACAGCTGCTGCTCTGASAGTTCCAAATGTTAAAATTAGACACCTCCAGCTACGACT	681
Db	10216	CACAGCTGCTGCTCTGASAGTTCCAAATGTTAAAATTAGACACCTCCAGCTACGACT	10157
OY	682	CAGGTGAGAGTTGCTGCTAAATAATTTCCCTCATCTCTTCTCACAGCAATGAATTTGCCA	741
Db	10156	CAGGTGAGAGTTGCTGCTAAATAATTTCCCTCATCTCTTCTCACAGCAATGAATTTGCCA	10097
OY	742	TCTGAACCCAAGTAAGAAAAAACAAATGGCTGCAATGTCTATATGATGATGCTACTACA	801
Db	10096	TCTGAACCCAAGTAAGAAAAAACAAATGGCTGCAATGTCTATATGATGATGCTACTACA	10037
OY	802	CAGATTTCTACGCTCTCCACAAAGTCAGAGATTTGAATGGTCAATACGACTTTTTT	861
Db	10036	CAGATTTCTACGCTCTCCACAAAGTCAGAGATTTGAATGGTCAATACGACTTTTTT	9977
OY	862	TTATTTCCCTTGACTCAGACAGCTACTCTTAATTTTGAAGACGTGTTTAAACCTTTGTGTG	921
Db	9976	TTATTTCCCTTGACTCAGACAGCTACTCTTAATTTTGAAGACGTGTTTAAACCTTTGTGTG	9917
OY	922	CTGGTTAAATTAATGTGTATACCTTTCTTCTCTCTGATACAGACGTTTCCCG	981
Db	9916	CTGGTTAAATTAATGTGTATACCTTTCTTCTCTCTGATACAGACGTTTCCCG	9857
OY	982	TGGTTGGTGTGAATTAATTTTGTGTAATGTTTAAATGTCATGTTAGATGTCAGGTTT	1041
Db	9856	TGGTTGGTGTGAATTAATTTTGTGTAATGTTTAAATGTCATGTTAGATGTCAGGTTT	9797
OY	1042	AOTCTCTGSAAGTAAGTCAACCATTCTTTATCAAAAGCAGACAGTCTGCTAC	1101
Db	9796	AOTCTCTGSAAGTAAGTCAACCATTCTTTATCAAAAGCAGACAGTCTGCTAC	9737
OY	1102	TTTCCAGCATTAATTTACTGAGATGTTAATGATGAAGTCTGATTTGGTACTTCTCC	1161
Db	9736	TTTCCAGCATTAATTTACTGAGATGTTAATGATGAAGTCTGATTTGGTACTTCTCC	9677

QY	1162	TGTAGAGTTAATAAAGGAAGATTACACTATCTGATTAATGTTCTTCATACCTGCAT	1221
Db	9616	TGTAGAGTTAATAAAGGAAGATTACACTATCTGATTAATGTTCTTCATACCTGCAT	9617
QY	1222	ATAATTTTGCGCTGAGAAATTTGTAATTTGTGCACACTATGTAAACAAACAACTGAAG	1281
Db	9616	ATAATTTTGCGCTGAGAAATTTGTAATTTGTGCACACTATGTAAACAAACAACTGAAG	9557
QY	1282	ATACTTTTAATTAATTAATTTGCTACTTATTTGGAAGTAAAAAAA	1323
Db	9556	ATACTTTTAATTAATTAATTTGCTACTTATTTGGAAGTAAAAAAA	9515
RESULT 9			
LOCUS	AL606522	203071 bp	DNA
DEFINITION	Mouse DNA sequence from clone RP23-19L22 on chromosome 11, complete sequence.		
ACCESSION	AL606522		
VERSION	AL606522.6	GI:20792609	
KEYWORDS	HTG.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 203071)		
TITLE	Almeida, J.		
AUTHORS	Direct Submission		
JOURNAL	Submitted (22-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humbry@sanger.ac.uk		
COMMENT	During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.		
	This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at <a href="http://www.sanger.ac.uk/Projects/C_elegans/wormpep">http://www.sanger.ac.uk/Projects/C_elegans/wormpep</a>		
FEATURES	from the RP23-19L22 is constructed by the group of Pieter de Jong. For further details see <a href="http://www.chori.org/bacpac/home.htm">http://www.chori.org/bacpac/home.htm</a>		
SOURCE	VECTOR: PBAC3.6.		
	location/Qualifiers		
	1. 203071		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/chromosome="11"		
	/clone="RP23-19L22"		
	/clone_11b="RP23-19L22"		
	/clone_11b="RP23-19L22"		
BASE COUNT	57371 a 44818 c 43421 g 57461 t		
ORIGIN			
Query Match	52.8%; Score 741.4; DB 10; Length 203071;		
Best Local Similarity	96.2%; Pred. No. 3,5e-110;		
Matches	792; Conservative 0; Mismatches 26; Indels 5; Gaps 3;		
QY	502	AGGAATTTGCTGATTCCTTGGAAATTCGTTTGGAAACAGTCTTAAGATGCAACGA	561
Db	177379	AGGAATTTGCTGATTCCTTGGAAATTCGTTTGGAAACAGTCTTAAGATGCAACGA	177438
QY	562	ATGTGAACACGCTTTCATGACAGTGCAGCTGAGATTAAAAAGCAATGGTCCCGGAG	621

Db 177439	ATGTTGAACAGCTCTTTTCATGACGATGCGACGTGAGATTAAACGGAATGGGTCCTGGAG	177498
QY 622	CAACAGCTGGTGCTCTGAGAACTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGCACT	681
Db 177499	CTACAGCTGGTGCTGCGCAGAGATCTCAATGTTAAATTCAGAGCACTCCAGTCAAGCACT	177558
QY 682	CAGGTGGAGGTGGCGCTAAATTTGGCTGCATCTCTTTCCATCAGCAATGAATTTGCAA	741
Db 177559	CAGGTGGAGGTGGCGCTAAATTTGGCTGCATCTCTTTCCATCAGCAATGAATTTGCAA	177618
QY 742	TCTGAACCCAGTGAAAAAACAATTCGCCGAATTTGACTGTATGTAGTCGCACTACAA	801
Db 177619	TCTGAACCCAGTGAAAAAACAATTTGCCGAATTTGACTGTATGTAGTCGCACTACAA	177678
QY 802	CAGATTTCTTACCGTCTCCACAAAGCTCAGAGATGTGAATGTGTCAATCTGAC--TTTTT	860
Db 177679	CAGATTTCTTACCGTCTCCAC--AAGCTCAGAGATGTGAATGTGTCAATCTGACCTTTT	177737
QY 861	TTTATTCCTTGACATCAGACAGCACTTTCATTTTCGAACTGTTTAAACCTTTGGT	920
Db 177738	TTTATTCCTTGACATCAGACCGGCTTCAATTTTCGAACTGTTTAAACCTTTGGT	177797
QY 921	GCTGGTTATTAATAATATGTGTGTATCTCTTGCTCTTCCGTATACAGACTGTTTCCC	980
Db 177798	GCTGGTTATTAATAATATGTGTGTATCTCTTGCTCTTCCGTATACAGACTGTTTCCC	177857
QY 981	GTCGTTGTTAGAAATATATTTTGGTTGATGCTTTATATTTGCAATGTTTATGTCAGGTT	1040
Db 177858	GTCGTTGTTAGAAATATATTTTGGTTGATGCTTTATATTTGCAATGTTTATGTCAGGTT	177917
QY 1041	TAGTCTTCTGAAAGATGAGTTCAGCACTTTTGTATCAACAACGACACAGTCTGTGCA	1100
Db 177918	TAGTCTTCTGAAAGATGAGTTCAGCACTTTTGTATCAACAACGACACAGTCTGTGCA	177977
QY 1101	CTTTCATGCAATTAAGTTTATGAGATGTTATATTAAGATCTGATTTGCTAGTCTTCC	1160
Db 177978	CTTTCATGCAATTAAGTTTATGAGATGTTATATTAAGATCTGATTTGCTAGTCTTCC	178037
QY 1161	TTGTAGAGTTATTAATGGAAGATATACATATCTGATTAAATAGTTTCTTCATCTGCA	1220
Db 178038	TTGTAGAGTTATTAATGGAAGATATACATATCTGATTAAATAGTTTCTTCATCTGCA	178097
QY 1221	TATATTTGTCGCGCACAATATTTGATTTTGTGCAACATGTAACAAACAACTGAA	1280
Db 178098	TATATTTGTCGCGCACAATATTTGATTTTGTGCAACATGTAACAA--AACTGAA	178154
QY 1281	GATATGTTTAATTAATATTTGACTATTTGGAAGTAAAAAAA 1323	
Db 178155	GATATGTTTAATTAATATTTGACTATTTGGAAGTAAAA 178197	
RESULT 10		
AL512349/c	116614 bp DNA linear HTG 10-JUL-2001	
LOCUS	Homo sapiens chromosome 6 clone RP11-223J2, *** SEQUENCING IN	
DEFINITION	PROGRES ***. In ordered pieces.	
ACCESSION	AL512349	
VERSION	AL512349.7 GI:13121473	
KEYWORDS	HTG; HTGS_PHAZE2; HTGS_CANCELLED.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	Bates, K.	
JOURNAL	Direct Submission	
COMMENT	Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk On Feb 23, 2001 this sequence version replaced gi:12956981. ----- Genome Center Center: Sanger Centre	





ACCESSION AL049545  
VERSION AL049545.6 GI:5002650  
KEYWORDS HMG; 60S Ribosomal Protein L7; RAB1; RAS; RPL7.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 104081)  
AUTHORS Smith,S.  
TITLE Direct Submission  
JOURNAL Submitted (06-JUL-1999) Sanger Centre, Hinxton, Cambridgeshire, UK  
COMMENT Submitted (06-JUL-1999) Sanger Centre, Hinxton, Cambridgeshire, UK  
On Jun 7, 1999 this sequence version replaced gi:4835284.  
During difference assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL  
This sequence is the entire insert of clone 263J7. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.  
This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>  
263J7 is from the library RPCI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCRPAC2>.  
FEATURES  
source  
1. 104081  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="q14.3-15"  
/clone="RPI-263J7"  
/clone\_11b="RPCI-1"  
1. .4880  
/note="LIP5 repeat: matches -237. .5623 of consensus"  
4895. .5820  
/note="LIP5 repeat: matches -1405. .-405 of consensus"  
5861. .6011  
/note="MLTJ repeat: matches 36. .174 of consensus"  
6032. .6086  
/note="L1 repeat: matches 3626. .3680 of consensus"  
6740. .8505  
/note="LIP12 repeat: matches 4202. .6165 of consensus"  
8512. .9328  
/note="LIP repeat: matches 3049. .3866 of consensus"  
9329. .9626  
/note="AluYb8 repeat: matches 1. .317 of consensus"  
9627. .10966  
/note="LIP repeat: matches 1712. .3049 of consensus"  
complement(11030. .11749)  
/gene="dJ2637.1"  
complement(11030. .11749)  
/gene="dJ2637.1"  
/gene="dJ2637.1"  
/note="dJ2637.1 (RPL7 (60S Ribosomal Protein L7) pseudogene): match: cDNAs: Em:X52967 Em:X57959 Em:L16558 Em:X57958 Em:M85235 Em:M17422 Em:M29016 Em:X57961 Em:X57960 Em:U63785; match: ESTs: Em:AI047212 Em:AI683841 Em:AI609411 Em:AI620251 Em:AI678713 Em:AA853290 Em:AI188448 Em:AI246632 Em:F28697.1 Em:AA640941 Em:AA147085 Em:AI354542 Em:AA648558 Em:AA172041

Em:AA736700 Em:AA181027 Em:AA617963 Em:AA129663 Em:AI385514 Em:AA58684 Em:AA148321 Em:AA147910 Em:AA0655193 Em:AA223084 Em:AA065240 Em:AA147610 Em:AI498784 Em:AA313779 Em:AA626741 Em:AA397390 Em:AA639843 Em:AA598466 Em:AA587875 Em:AA314579 Em:AA081412 Em:AA838601 Em:AA129729 Em:AA236303 Em:AA580325 Em:AA147597 Em:AA313220 Em:AA516160 Em:AA508128 Em:AA562608 Em:AA635115 Em:AI367854 Em:AA607384 Em:AA846920 Em:AA608110 Em:AA745100 Em:AA607384 Em:AA507813 Em:AA314838 Em:AA736534 Em:AA591026 Em:AA008632 Em:W77465 Em:AA648073 Em:AA648202 Em:AA552160 Em:AA215370 Em:DS1697 Em:DS1979 Em:DS5053 Em:AA760175 Em:AA856854 Em:N98334 Em:AA649499 Em:AA879164 Em:AA332842 Em:AA88688 Em:AA173529 Em:AA091272 Em:AA093750; match: proteins: Sw:P18124 Sw:P14148 Sw:P05426 Sw:Q42208 Sw:O01802 Sw:O60143 Sw:P25457 Sw:P1874 Sw:P32100 Sw:P05737 Sw:Q12213 Tr:O14371 Tr:O08953"  
/pseudo  
/codon\_start=-1  
/evidence=not\_experimental  
11307. .11789  
/note="match: GSS A081091 B54651 A0028779"  
11790. .12860  
/note="LIP13 repeat: matches 647. .1730 of consensus"  
12908. .13190  
/note="AluJo repeat: matches 3. .302 of consensus"  
13341. .13560  
/note="HAI1 repeat: matches 1483. .1721 of consensus"  
13346. .13347  
/note="3 subclones read TCTT, 2 read TCT"  
13594. .15880  
/note="LIPB2 repeat: matches 3950. .6154 of consensus"  
15881. .16608  
/note="LIPB4 repeat: matches 5413. .6144 of consensus"  
16609. .19090  
/note="LIPB2 repeat: matches 1394. .3950 of consensus"  
19096. .19323  
/note="LIP3 repeat: matches 3. .231 of consensus"  
19472. .23244  
/note="LIP repeat: matches 638. .4389 of consensus"  
23240. .23462  
/note="LIPBa repeat: matches -1549. .-1330 of consensus"  
23577. .29885  
/note="LIPB3 repeat: matches -218. .6130 of consensus"  
29886. .29943  
/note="LIPAS repeat: matches 6081. .6139 of consensus"  
29944. .30234  
/note="LIPB3 repeat: matches -508. .-218 of consensus"  
complement(30558. .30912)  
/note="match: EST AA528504"  
31385. .31601  
/note="MIR repeat: matches 29. .256 of consensus"  
31721. .32002  
/note="LIPB7 repeat: matches 5862. .6145 of consensus"  
32279. .32583  
/note="AluSx repeat: matches 1. .308 of consensus"  
33356. .33592  
/note="MIR repeat: matches 8. .247 of consensus"  
33779. .37507  
/note="LIM2 repeat: matches 2533. .6308 of consensus"  
37508. .37811  
/note="AluY repeat: matches 3. .305 of consensus"  
37812. .37985  
/note="LIM2 repeat: matches 2362. .2533 of consensus"  
37976. .41630  
/note="LIM1 repeat: matches -1390. .2698 of consensus"  
41878. .41924  
/note="MER5B repeat: matches 127. .173 of consensus"  
42325. .42418  
/note="MDE1 repeat: matches 1. .79 of consensus"  
42451. .42853  
/note="MLTIC repeat: matches 23. .466 of consensus"

```
misc.feature      complement(42510:.42919)
/note="match: STR 279267"
42912..43060
repeat_region    /note="L2 repeat: matches 2205..2346 of consensus"
45024..45322
/note="AlusP repeat: matches 1..300 of consensus"
45665..4616
/note="WSTA repeat: matches 1..426 of consensus"
46284..46709
/note="WSTA repeat: matches 1..426 of consensus"
47275..47377
/note="L2 repeat: matches 2604..2710 of consensus"
47897..48335
/note="match: GSS AQ121187"
48336..48736
/note="WLRII repeat: matches 1..390 of consensus"
49379..49625
/note="WIR repeat: matches 3..255 of consensus"
49802..50099
/note="L2 repeat: matches 1171..1480 of consensus"
50099..52068
/note="L2 repeat: matches 2683..2748 of consensus"
50164..51146
/note="L2 repeat: matches 2247..2498 of consensus"
50750..51146
/note="L2 repeat: matches 14..432 of consensus"
51505..51706
/note="L2 repeat: matches 2544..2745 of consensus"
52046..52185
/note="WIR repeat: matches 48..196 of consensus"
52644..52866
/note="charlie5 repeat: matches 2364..2585 of consensus"
53590..55137
/note="L1 repeat: matches -665..1099 of consensus"
55150..57986
/note="L1 repeat: matches 3281..6133 of consensus"
57988..58374
/note="L1 repeat: matches 1087..1477 of consensus"
58375..58932
/note="L1 repeat: matches 1..553 of consensus"
58933..59661
/note="L1 repeat: matches 1477..2570 of consensus"
59661..59796
/note="L1 repeat: matches 1477..2570 of consensus"
59796..59796
/note="FAM repeat: matches 61..156 of consensus"
59802..59827
/note="L1 repeats 2 mer ac 92% conserved"
59829..60191
/note="L1 repeat: matches 2181..2545 of consensus"
60280..60605
/note="AlusX repeat: matches 1..309 of consensus"

Query Match      48.8%; Score 686; DB 9; Length 104081;
Best Local Similarity 74.8%; Pred. No. 3e-101; Indels 150; Gaps 13.
Matches 1098; Conservative 0; Mismatches 220;
```

D	95576	AGTTTGGAGTGGAAACATGTATATACATATACAGAAACATCAATCAGCAATTAAGTGG	95517
OY	318	GATTTCAAAATAGAACTCTAGAGTTAGAGGGGAAACAAATCAAGCTTAAATA-----	371
D	95516	AATTTCAAAATAGAACTATATATATATAGCAAAATATATCAAGCTCACTAGGGAC	95457
OY	372	-----	371
D	95456	ACAACAGCCAGAAAATTTTCATCACTCCAGTTATATACAGAGGCTAAGGCTGTA	95397
OY	372	-----GATCTCCATATATATGTTTAAACAGGCGCTGAGAAATATGATC	415
D	95396	TGATATGACAAATCAGAGTCTCTTGATATATATGTCAAATATGTTGGTAGGAAATATGATCA	95337
OY	416	TTATGCCAGTGAATAATGTGACAAATTTGGTAGGGGAAACAAATGATCATACACAAA	475
D	95336	CTAGGCACTGAAATATGTCTACAAATTTGTTGTAGGAAACAAATGTAT--TGCACAAA	95278
OY	476	GAATAGTATGATGATGACAAACAGGAGAAATTTGCTATTCCTTGGAATTCCTGTTTT	535
D	95277	GAA--AGTAACTTACCAACAAACAAAGATTTGCTATTCCTTGGAATTCCTGTTTT	95221
OY	536	GGAACCACTGCTAAAGTGCACAAATGTATGAAACAGCTTTTCATGACGATGGACATGA	595
D	95220	GGAATCTGTGTATAGACACAGCAAAATGTATGAAACAGATTTTCATGAAATGGCATGGA	95161
OY	596	GATTAATAAGCAATGGCTGCCGGAGCAACAGCTGTGGCTGTGAAATCCAAATGTAA	655
D	95160	GATTAATAAGCAATGGG-ACATAGACAAACAGCTGTGTATGACAGAAAGCTAATTTTAA	95102
OY	656	AATTCAGACATCTCAAGTGAAGATGACAGGATGAGGTGCTCTAAATTTGGCTCATC	715
D	95101	AATTTAGACACTTTTGTCTACAGATGACAGTGAAGTGTATCTAAATTTGGCTCATC	95042
OY	716	CTTTTCTCACACATGAATTTGCATCTGAAACCAAGTGAATAACAAATTTGCTGAA	775
D	95041	CTTTTCTCACAA-CATTAATTTGCATCTTAACCCAAATGAAAAACAAATTTGCTGAA	94983
OY	776	TTTACTGATGTACTGCACTCAACAGATTTCTACCCTGCTCCAAAGCTAGATG	835
D	94982	TTTACTGATGTACTGCACTCAACAGATTTCTACCCTGCTCCAAAGCTAGATG	94923
OY	836	GTAATGTGCTATCTACTCTTTTCTTTTCTTCCCTGACACGAAGCTCACTCATTT	895
D	94922	GCAATATGCTATCTGCA-ATTTTTTTATACCTTACACCAAGATACTTAATTCAGTT	94864
OY	896	TCAGACTGTTTAAACCTTGTGTGCTGCTGTTTAAATAATGTGTATCTCTGGTG	955
D	94863	TCAGAAACCTTTAAACTTTGTGTGCTGCTTACAAATACTGCTATCTCTTGCA	94804
OY	956	CTTTCTGCATACAGACGTTCCTCCGGTGGTT-----AGAAATATTTTGGTTT	1007
D	94803	CTTTCTGCATACAGATCTAGCTCCCATGATTTCTTTGTAATTTAGATTTATATTTT	94744
OY	1008	GATGTTATATGGCACTGTTATATGATGCAAGTTTATAGCTTCTCAATAATGATTCAGCA	1067
D	94743	GATGTTATATGGCACTGTTATATGATGCAAGTTTATAGCTTCTCAATAATGATTCAGCA	94684
OY	1068	TTTGTATCAACAGCAAGCAAGCTGTGCACTTCCATCATTAAGTTTGTGAGA	1126
D	94683	CTTTGTATCAACAGCAAGCAAGCTGTGCA-----TTCATGTATACAG	94636
OY	1127	TGTATATATGAACATGTATGTTGCTATCTCTCTGTAGAGTTATTAATGGAAAGTTA	1186
D	94635	TGTATATATGAACATGTATGTTGCTATCTCTCTGTAGAGTTATTAATGGAAAGTTA	94577
OY	1187	CACATCTGATTAATGTTCTCTCATCTCTGATATATATTTGGCTGCAAGATATATG	1246
D	94576	TACATTTATTAATGTTCTCTCATCTCTGATATATATTTGGCTGCAAGATATATG	94517
OY	1247	AATTTGTCACACTATGATTAACAAACAACTGAAATGTTTAAATTAATTTGACTTA	1306
D	94516	GATTTGTCACACTATGAT-ACAAACAACTAAGATGTTTAAACAAATATGACTTA	94458



repeat\_region /rpt\_family="AT\_rich"  
 17221..17290 /rpt\_family="L2"  
 repeat\_region 17828..17963 /rpt\_family="L2"  
 repeat\_region 17864..18138 /rpt\_family="MERL\_type"  
 repeat\_region 18139..18209 /rpt\_family="L2"  
 repeat\_region 18210..18521 /rpt\_family="Alu"  
 repeat\_region 18522..18865 /rpt\_family="L2"  
 repeat\_region 19330..19710 /rpt\_family="L2"  
 repeat\_region 20271..20291 /rpt\_family="AT\_rich"  
 repeat\_region 21212..21327 /rpt\_family="MIR"  
 repeat\_region 21523..21828 /rpt\_family="Alu"  
 repeat\_region 22806..22860 /rpt\_family="CT-rich"  
 repeat\_region 22994..23105 /rpt\_family="MIR"  
 repeat\_region 25554..25846 /rpt\_family="Alu"  
 repeat\_region 25827..26475 /rpt\_family="L2"  
 repeat\_region 26500..26664 /rpt\_family="L2"  
 repeat\_region 26754..26897 /rpt\_family="MERL\_type"  
 repeat\_region 26898..27113 /rpt\_family="L2"  
 repeat\_region 27513..27652 /rpt\_family="L2"  
 repeat\_region 28002..28057 /rpt\_family="MERL\_type"  
 repeat\_region 28065..28196 /rpt\_family="Alu"  
 repeat\_region 28481..28541 /rpt\_family="MERL103"  
 repeat\_region 30093..30138 /rpt\_family="TG)n"  
 repeat\_region 33232..33532 /rpt\_family="Alu"  
 repeat\_region 36502..36816 /rpt\_family="Alu"  
 repeat\_region 37354..37634 /rpt\_family="Alu"  
 repeat\_region 41942..41974 /rpt\_family="L2"  
 repeat\_region 42371..42439 /rpt\_family="MIR"  
 repeat\_region 42841..42718 /rpt\_family="TA)n"  
 repeat\_region 42719..42746 /rpt\_family="CA)n"  
 repeat\_region 43845..43966 /rpt\_family="L2"  
 repeat\_region 47469..47686 /rpt\_family="MERL\_type"  
 repeat\_region 47727..47789 /rpt\_family="L2"  
 repeat\_region 47840..48153 /rpt\_family="L2"  
 Query Match 48.7%; Score 684.4; DB 9; Length 162471;  
 Best Local Similarity 71.1%; Pred. No. 5,4e-101;  
 Matches 1237; Conservative 0; Mismatches 76; Indels 427; Gaps 8;

Db 69889 AAGCAATAGCTGATGCGGCGGCTGATTCCTTCAGGGGCGAGAGTGGGGAAGA 69830  
 QY 60 GCTTGGCTCCCGGGAAGCCCTATTCATTCCTTCTTCGATTACCGGTGGCCGAG 119  
 Db 69829 CATTTACTCCAGAACGCCCTATCTCATTCCTTCATGATACCGGTGTCAGAG 69770  
 QY 120 ACTCAGGGCGCGCTGCGGAGAGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGG 179  
 Db 69769 ACTCAGGGCGGAGCTGTAGCAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 69710  
 QY 180 ATGTCAGCATGATCCCGAATATGATTTATTCAGTTACTTCTGATTGGGACATCA 239  
 Db 69709 ATGTCAGCATGATCCCGAATATGATTTATTCAGTTACTTCTGATTGGGACATCT 69651  
 QY 240 GGGGTTGGAAGTCTTCT 298  
 Db 69650 GGGGTTGGAAGTCTTCT 69591  
 QY 299 CATCAGACAAATGGTGTGATTTCAAAATTAAGACTATAGATTAGAGGGAACAAAT 358  
 Db 69590 CATCAGACAAATGGTGTGATTTCAAAATTAAGACTATAGATTAGAGGGAACAAAT 69531  
 QY 359 CAAGCTCAAAATG----- 371  
 Db 69530 CAAGCTCAAAATGGAACAGACAGAACAGAAAGTTGGAACAGTATCTCCAGTTA 69471  
 QY 372 -----GAGTCTTGA 382  
 Db 69470 TTACAGAGGAGCCATGATCTCTAGTGTATGATGTGACAGATCAGAGCTTGA 69411  
 QY 383 TAATGTAAACAGTGTGAGAGAAATATGCTTATGCGAGTGAATATGCAAAATTT 442  
 Db 69410 TAATGTAAACAGTGTGAGAGAAATATGCTTATGCGAGTGAATATGCAAAATTT 69351  
 QY 443 GTTGATGGAAGCAAAATGTATGTGACCAAGAAAGTATGATGACCAAGAGCA 502  
 Db 69350 GTTGATGGAAGCAAAATGTATGTGACCAAGAAAGTATGATGACCAAGAGCA 69291  
 QY 503 GGAATTTGATTCCTTCTGGAATTCCTTCTGGAATTCCTTCTGGAATTCCTTCTGGA 548  
 Db 69290 GGAATTTGATTCCTTCTGGAATTCCTTCTGGAATTCCTTCTGGAATTCCTTCTGGA 69231  
 QY 549 ----- 548  
 Db 69230 GCGCGAGCGGTGCTCATGCTGTATTCACAGCACTTTGGAGCCGAGGTGGTGA 69171  
 QY 549 ----- 548  
 Db 69170 CACCTGAGGTGAGAGTTGAGAGACAGCTGGCAATGATGAAACCCATCCTACTA 69111  
 QY 549 ----- 548  
 Db 69110 AAAATACAAAATTAGCCAGGCAATGTGGCGGCGCTGTATCCAGCTACTTGGAG 69051  
 QY 549 ----- 548  
 Db 69050 CTGAGGAGAGGAATTCGTAACCCAGGTGGGAGGTTCAGTGAGCGAGATTGGCC 68991  
 QY 549 -----AAG 551  
 Db 68990 CATTTCTCTCCAGCTAGAGACGAGCAAGACTCTGTCTCAAAAAAAGAAAG 68931  
 QY 552 AATCCAGCAATGTAGAACAGTCTTTCATGATGAGCAGCTGAGATTAAAAAGGAATG 611  
 Db 68930 AATCCAGCAATGTAGAACAGTCTTTCATGATGAGCAGCTGAGATTAAAAAGGAATG 68871  
 QY 612 GGTCCGAGCAAGCTGT 671  
 Db 68870 GGTCCGAGCAAGCTGT 68811  
 QY 672 GTCAAGAGTCAAGGTGAGGT 731  
 Db 68810 GTCAAGAGTCAAGGTGAGGT 68751

QY 732 GAATTTGCAATCTGAACCAAGTGAAAAACAAATTCCTGAATTGACTGTATGTAGC 791  
|||||  
Db 68750 AAATTTGCAATCTGAACCAAGTGAAAAACAAATTCCTGAATTGACTGTATGTAGC 68691  
QY 792 TGCACCTACAAAGATCTTACCCTCTCCACAAGGTGAGATGTGTAATGTGCTATCT 851  
|||||  
Db 68690 TGCACCTACAAAGATCTTACCCTATATCCACAGAGGTGAGATGTGTAATGTGCTATCT 68631  
QY 852 GAC--TTTTTTTATTCCTTGACTCAGACAGCTACTTCATTTTCAGAACCTGTTTA 909  
|||||  
Db 68660 GACTTTTTTTTTTATTTCTCTGTGACTCAAGACAGCTAATCTTTCAGAACCTGTTTA 68571  
QY 910 AACCTTGTGTGCTGTTTAAATAATGTTGTGTAATCCTGTGCTTCTCCTGATACCA 969  
Db 68570 AACCTTGTGTGCTGTTTAAATAATGTAAT-----TTTCCTGATATCA 68525  
QY 970 GACTG-TTTCCTCCGTGCTGTTAGATATATTTGTTTGTATGTTTATGTCATGTTT 1028  
|||||  
Db 68524 GACTGTTTCTCGTGGTGGTTAGATATATTTGTTTGTATGTTTATGTTGCTGTTT 68465  
QY 1029 AGATGTCAGTTTATGTTCTTGAAGATGAAGTTCAGCCATTTTGTATCAACAGCACAAG 1088  
|||||  
Db 68464 AGATGTCAGTTTATGTTCTTGAAGATGAAGTTCAGCCATTTTATATCAACAGCACAAG 68405  
QY 1089 CAGTGTCTGACCTTCCATGCAATGAAGTTAGTGAATGTTAATGTAAGATCTGATTT 1148  
Db 68404 CAGTGTCTGACCTTCCATGCAATGAAGTTAGTGAATGTTAATGTAAGATCTGATTT 68345  
QY 1149 GCTAGTCTCTCTCTGTAGAGTATTAATGAAGATTCACATCTGTTTAAATGTTCTT 1208  
|||||  
Db 68344 GCTAGTCTCTCTCTGTAGAGTATTAATGAAGATTCACATCTGTTTAAATGTTCTT 68285  
QY 1209 TCATACCTGTCATTAATTTGCTGCAGATATTTGAATTTGTGCACACTATGTAAC 1268  
Db 68284 TCATACCTGTCATTAATTTGCTGCAGATATTTGAATTTGTGTGATCTATGTAAC 68225  
QY 1269 AAAACAACCTGAAGATGTTTAAATATATGTAATTTGTAAGTGAAGTAAAAA 1328  
Db 68224 AAAACAACCTGAAGATGTTTAAATATATGTAATTTGTAAGTGAAGTAAAAA 68165

RESULT 13  
MMPT14 4678 bp DNA linear ROD 05-APR-1995  
LOCUS MMPT14 4678 bp DNA linear ROD 05-APR-1995  
DEFINITION Mouse ypc1 gene for ras-related GTP-binding protein, exons 4-6.  
ACCESSION X15747  
VERSION X15747.1 GI:55463  
KEYWORDS GTP-binding protein; ras-related protein; ypc1 gene.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 4678)  
Wichmann, H.  
TITLE Direct Submission  
JOURNAL Submitted (07-JUL-1989) Wichmann H., MPI of Biophysical Chemistry,  
Dept of Molecular Genetics, P O Box 2841, D-3400 Goettingen, F R G  
2 (bases 1 to 4678)  
AUTHORS Wichmann, H., Disela, C., Haubruck, H. and Gallwitz, D.  
TITLE Nucleotide sequence of the mouse ypc1 gene encoding a ras-related  
GTP-binding protein  
JOURNAL Nucleic Acids Res. 17 (16), 6737-6738 (1989)  
MEDLINE 89386011  
PUBMED 2506528  
COMMENT Data kindly reviewed (27-OCT-1989) by Wichmann H.  
FEATURES  
source location/Qualifiers  
1..4678  
/organism="Mus musculus"  
/strain="c(CH)/c(CH)"  
/db\_xref="taxon:10090"  
precursor\_RNA <1..2682  
/note="Primary transcript"

exon 239..334  
/note="Exon 4"  
/usedin=X15744:gtfb\_cds  
Intron 335..1521  
/note="Intron IV"  
exon 1522..1653  
/note="Exon 5"  
/usedin=X15744:gtfb\_cds  
Intron 1654..1802  
/note="Intron V"  
mRNA 1803..2682  
/note="Exon 6"  
/note="2000  
/note="GTP-binding protein, exon 6 (AA 141-205)"  
/usedin=X15744:gtfb\_cds  
misc\_feature 2586..2591  
/note="pot. polyA signal (1.6 kb mRNA)"  
misc\_feature 2654..2659  
/note="pot. alt. polyA signal (1.6 kb mRNA)"  
polyA\_site 2682  
/note="polyA site"  
BASE COUNT 1338 a 913 c 905 g 1522 t  
ORIGIN  
Query Match 48.6%; Score 682.6; DB 10; Length 4678;  
Best Local Similarity 94.8%; Pred. No. 9.5e-101;  
Matches 782; Conservative 0; Mismatches 34; Indels 9; Gaps 7;  
QY 502 AGGAATTTGCGATTCCTTGGAAATTCGTTTGGAAACAGTGTAGAAATGCAACGA 561  
Db 1801 AGGAATTTGCGATTCCTTGGAAATTCATTTTGGAAACAGTGTAGAAACAGCA 1860  
QY 562 ATGTAGAACAGTCTTTCATGACGATGCGAGCTGAGATTAATAAGCAATGGTCCCGAG 621  
Db 1861 ATGTAGAACAGTCTTTCATGACGATGCGAGCTGAGATTAATAAGCAATGGTCCCGAG 1920  
QY 622 CAACAGCTGT 681  
Db 1921 CTACAGCTGT 1980  
QY 682 CAGGTGAGGT 741  
Db 1981 CAGGTGAGGT 2040  
QY 742 TCTGAACCAAGTGAAAAACAAATTCCTGAATTTGCTGTATGTATGATGCTACTACAA 801  
Db 2041 TCTGAACCAAGTGAAAAACAAATTCCTGAATTTGCTGTATGTATGATGCTACTACAA 2100  
QY 802 CAGATTTCTTACCGTCTCCACAAGGTGAGATTTGTAATGTGCAATCTGAC-TTTT 860  
Db 2101 CAGATTTCTTACCGTCTCCACAAGGTGAGATTTGTAATGTGCAATCTGAC-TTTT 2159  
QY 861 TTTATTCCTTGATGCAAGACGTAATCTTATTTTTCGAAC-TGTTTAAACCTTGTG 919  
Db 2160 TTTATTCCTTGATGCAAGACGTAATCTTATTTTTCGAAC-TGTTTAAACCTTGTG 2219  
QY 920 TGTGTTTAAATAATATGTGTATCTTGTGCTTCTTCTGATACCAACAGCTTTC 979  
Db 2220 TGTGTTTAAATAATATGTGTATCTTGTGCTTCTTCTGATACCAACAGCTTTC 2279  
QY 980 CGT 1039  
Db 2280 CGT 2339  
QY 1040 TTAGTCTCTGAAGTGAAGTTCAGCCCTTTGTATCAACAGCAACAGCAGTGTCTG 1099  
Db 2340 TTAGTCTCTGAAGTGAAGTTCAGCCCTTTGTATCAACAGCAACAGCAGTGTCTG 2399  
QY 1100 ACTTTCATG-CATTAAGTTAGTGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 1158  
Db 2400 AGTTTCCAGTCATTAAGTTAGTGTATGTATGTATGTATGTATGTATGTATGTATGTATGT 2458  
QY 1159 CTTGTAGAGTTTAAATGAAGATTAACACTATCTGATTAATAGTTTCTTCACTCTG 1218

Db 2459 CCGGTAGAGTTATTAAGGACAGACACTATGATTAATATCTTCTTCACTCTG 2518  
 Oy 1219 CATTAATTTTGGTCGACAGATTTGATTTGGACACTATGTAACAACATCG 1278  
 Db 2519 CATTAATTTTGGTCGACAGATTTGATTTGTTGGACACTATGTAACA---AAGC 2574  
 Oy 1279 AAGATATGTTAATAATATGTTACTTATTTGGAAGTAAAAAAA 1323  
 Db 2575 AGATATTTTAAATTAATATGTTACTTATTTGGAAGTAAAA 2619

RESULT 14  
 AC117841 207945 bp DNA linear HTG 17-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-36006, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\*; 73 unordered pieces.  
 AC117841  
 VERSION AC117841.4 GI:21746140  
 KEYWORDS HTG: HTGS\_PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 207945)  
 Muny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Albrooks,S.L., Amaratunga,H.C., Are,J.R., Ayelo,M., Banks,T.,  
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonini,D.,  
 Bouck,J., Bowles,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Bulay,C., Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carrott,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Day-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Donthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,N., Gili,R.,  
 Galsi,A., Geo,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gortelli,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,  
 Hernandez,C., Harris,K., Hart,M., Haylak,P., Hayes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogue,M., Hollway,C., Hollins,B.,  
 Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Katsios,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Kirovic,J., Kishida,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisgud,H.,  
 Lozardo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,N., Nickerson,E., Nwokenkwo,S., Ogutu,M., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogutu,M., Nguyen,N.,  
 Oreguay,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pichens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rivers,M., Rojas,A., Rojuben,I., Rolfe,M., Ruiz,S., Saevery,G.,  
 Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,  
 Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,O.,  
 Umanu,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,O.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wlaczek,K., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstock,G. and Gdbbs.R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 207945)  
 REFERENCE  
 TITLE  
 AUTHORS  
 DIRECT SUBMISSION  
 JOURNAL  
 Submitted (11-Apr-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One

REFERENCE  
 BAYLOR PLAZA, HOUSTON, TX 77030, USA  
 3 (bases 1 to 207945)  
 TITLE  
 AUTHORS  
 DIRECT SUBMISSION  
 JOURNAL  
 Submitted (17-Jul-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:20258099.  
 COMMENT  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 Project Information  
 Center project name: GIBP  
 Center clone name: CH230-36006  
 Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 129522 bases at least Q40  
 Consensus quality: 134968 bases at least Q30  
 Consensus quality: 138682 bases at least Q20  
 \*\*\*\*\*  
 NOTE: Estimated insert size may differ from sequence length  
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 73 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N. The exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 1 1046: contig of 1046 bp in length  
 1047 1145: gap of unknown length  
 1147 2183: contig of 1037 bp in length  
 2184 2283: gap of unknown length  
 2284 3421: contig of 1138 bp in length  
 3422 3521: gap of unknown length  
 3522 4686: contig of 1165 bp in length  
 4687 4786: gap of unknown length  
 4787 6158: contig of 1372 bp in length  
 6159 7327: gap of unknown length  
 7328 7427: contig of 1069 bp in length  
 7428 9082: contig of 1655 bp in length  
 9083 9182: gap of unknown length  
 9183 10231: contig of 1049 bp in length  
 10232 10331: gap of unknown length  
 10332 11663: contig of 1332 bp in length  
 11664 11763: gap of unknown length  
 11764 13476: contig of 1713 bp in length  
 13477 13576: gap of unknown length  
 13577 15050: contig of 1474 bp in length  
 15051 15150: gap of unknown length  
 15151 16522: contig of 1372 bp in length  
 16523 18019: gap of unknown length  
 18020 18119: contig of 1397 bp in length  
 18120 19774: gap of unknown length  
 19775 21203: contig of 1655 bp in length  
 21204 21303: gap of unknown length  
 21304 22991: contig of 1329 bp in length  
 22992 23091: contig of 1688 bp in length  
 23092 24721: gap of unknown length  
 24722 24821: gap of unknown length  
 24822 25833: gap of 1012 bp in length  
 25834 27588: gap of unknown length  
 27589 27688: contig of 1635 bp in length  
 27689 29226: gap of unknown length  
 29227 29226: contig of 1538 bp in length

*	29327	29326:	gap of unknown length
*	29327	30424:	contig of 1098 bp in length
*	30425	30524:	gap of unknown length
*	30525	32145:	contig of 1621 bp in length
*	32146	32245:	gap of unknown length
*	32246	33498:	contig of 1253 bp in length
*	33499	33598:	gap of unknown length
*	33599	34724:	contig of 1126 bp in length
*	34725	34824:	gap of unknown length
*	34825	36502:	contig of 1678 bp in length
*	36503	36602:	gap of unknown length
*	36603	37971:	contig of 1369 bp in length
*	37972	38071:	gap of unknown length
*	38072	39263:	contig of 1192 bp in length
*	39264	39363:	gap of unknown length
*	41129	41228:	contig of 1765 bp in length
*	41229	41229:	gap of unknown length
*	42551	42750:	contig of 1422 bp in length
*	42751	44150:	gap of unknown length
*	44151	44250:	contig of 1400 bp in length
*	44251	44240:	gap of unknown length
*	45441	45540:	contig of 1190 bp in length
*	45441	47375:	gap of unknown length
*	47376	47475:	contig of 1835 bp in length
*	47476	48634:	gap of unknown length
*	48635	48734:	contig of 1159 bp in length
*	48735	50080:	gap of unknown length
*	50081	50180:	contig of 1346 bp in length
*	50181	51471:	gap of unknown length
*	51472	51571:	contig of 1291 bp in length
*	51572	52959:	gap of unknown length
*	52960	53059:	contig of 1388 bp in length
*	53060	54842:	gap of unknown length
*	54843	54942:	contig of 1783 bp in length
*	54943	57384:	gap of unknown length
*	57385	57484:	contig of 2442 bp in length
*	57485	57485:	gap of unknown length
*	59404	59403:	contig of 1919 bp in length
*	59504	60963:	gap of unknown length
*	60964	61063:	contig of 1460 bp in length
*	61064	62330:	gap of unknown length
*	62331	62430:	contig of 1267 bp in length
*	62431	63700:	gap of unknown length
*	63701	63800:	contig of 1270 bp in length
*	63801	65110:	gap of unknown length
*	65111	65210:	contig of 1310 bp in length
*	65211	66855:	gap of unknown length
*	66856	66995:	contig of 1685 bp in length
*	66996	68391:	gap of unknown length
*	68392	68491:	contig of 1396 bp in length
*	68492	70672:	gap of unknown length
*	70673	70772:	contig of 2181 bp in length
*	70773	72373:	gap of unknown length
*	72374	72473:	contig of 1601 bp in length
*	72474	75328:	gap of unknown length
*	75329	75428:	contig of 2855 bp in length
*	75429	77871:	gap of unknown length
*	77872	77971:	contig of 2443 bp in length
*	77972	80743:	gap of unknown length
*	80744	80843:	contig of 2772 bp in length
*	80844	82486:	gap of unknown length
*	82487	82586:	contig of 1643 bp in length
*	82587	84732:	gap of unknown length
*	84733	84832:	contig of 2116 bp in length
*	84833	87219:	gap of unknown length
*	87220	87319:	contig of 2387 bp in length
*	87320	90363:	gap of unknown length
*	87320	90363:	contig of 3044 bp in length

Query Match

Best Local Similarity

Matches 937; Conservative

48.3%;

81.9%;

0;

Score 679.2;

Pred. No. 3.8e-100;

Mismatches 98;

DB 2;

Indels 109;

Gaps 9;

Length 207945;

QY	1	AAGCGATAGCTAGT- GCGCGCGCTGCTCATTTGTGTTCTTAGGGGACGGAGTATAGGGAA	59
D	91854	AAGCATACACTAGTGGTGGCGGCTGCTATTTGTATTAAGCGTGGTGT- GGGGTTTA	91796
QY	60	CGTTTGCCTCCGGAAACGCGCATCTCATTCCTTTCTTGCATTCACCGTGGCGGGAG	119
D	91795	CGTTTGTCTACTGGAAACGCTTAGCTTCTCTCCCTTCCATTATACCGTGGTGCAGAA	91730
QY	120	AGTCAGGGCGGCGGCTGCGGCAGCAAGGGCGGGCGGCGCGGCAGCTGCACTGAC	179
D	91735	GCTTGGGGGGGAGCGTGGCGCTCAGAAAG-- GGGGGTGGCAGGGGAGCGAGCTGCAGTAC	91679
QY	180	ATGTCAGCATTAATCCGAAATATGATTTATTTCAAGTACTTGTATTTGGCGACTA	239
D	91678	ATGTCAGCATTAATCCGAAATATGATTTATTTCAAGTACTTGTATTTGGCGACTA	91619
QY	240	GGGGTTGGAAACGTGGCTCTCTCTCTAGTTTGGAGATATATATATATACAGAACTAC	299
D	91618	GGGGTTGGAAAGCTGTGCTTCTCTCTCTAGTTTGGAGATATATATATATACAGAACTAC	91556
QY	300	ATCAGCACAA- TTGTTGTGATTTTCAAAATTAAGAACTATATAGATTATAGAGGAAAAAT	358
D	91558	ATCAGCATATTTTGGTGTGATTTCAAGATATAGAGACTATATAGAGCTTGAAGGAAAAATAT	91499
QY	359	CAAGCTTCAATA- -----	371
D	91498	CAAGCTTCAAGATATGGGACACAGACAGCGCCAGAAAGATTTGCAACATCACTCCAGTTA	91430
QY	372	----- GAGTCCTTCA	382
D	91438	TTACAGAGAGAACCATTTATCATATAGTTGTATGTATGTACCGACCGAGAGTCTTCAA	91378
QY	383	TATGTTTAAACAGTGGCTGCAGAAATAGATCGTTATGCGAGTGAATAATGCAACAAAT	442
D	91378	TATGTTTAAACAGTGGCTGCAGAAATAGATCGTTATGCGAGTGAATAATGCAACAAAT	91315
QY	443	GTTGTAGGAAACAAATGTATGTACCAACAAGAAATATAGTACTACACAACAGCGAA	502
D	91318	GTTGTAGGAAAGAACACATGTGACTCCACAAC- - - AAGATATGTTATGATACACAACAGCAAA	91263
QY	503	GGAATTTGCTGATTCCTCTGGAATTCGTTTTGGAACACAGTGTAAACATATGCAACGAA	562
D	91262	GGAATTTGAGATTCCTCTGGAATTCATTTTGGAAACAGTGTAAACACAAATGAT	91203
QY	563	TGTAGAACAGTCTTTTCATCAGATGGACGACTGATTTAAAGAACGAATGGTCCCGAGC	622
D	91202	TGTAGAACAGTCTTTTCATCAGATGGACGACTGATTTAAAGAACGAATGGTCCCGAGC	91144
QY	623	AACAGCTGTGTGCTGAGAACTCCAAATGTAAATTCAGAGCACTCCAGTCAAGCAATC	682
D	91142	AACAGCTGTGTGCTGAGAACTCCAAATGTAAATTCAGAGCACTCCAGTCAAGCAATC	91083
QY	683	AGGTGAGGTGCTGCTAAATTTGCTCTCATCTTTTCTCACAGCAATGAATTTGCAAT	742
D	91082	AGGTGAGGTGCTGCTAAATTCGTCAACGTCCTTTTCTCACAGCAATGAATTTGCAAT	91023
QY	743	CTGAACCCCAAGGAAAAAACAAT- TGCTGATATGTACTGATGTACTGACTGACATCA	801
D	91022	CTGAACCCCAAGGAAAAAACAATGTGCTGATATGTACTGATGTACTGACTGACATCA	90963
QY	802	CAGATTTCTTACCGTCTCCCAAGGTCAGAGATTTGTAATGTGCAATACTGAC- TTTT	860
D	90962	CAGATTTCTTACCGTTTTCAC- AAGTCAGAGATTTGTAATGTATGATCAATACTGACTTTT	90904
QY	861	TTTTTCCCTGACTCAGACAGCTATCTTATTTTTCAGAACTGTTTAAACCTTTGTGT	920
D	90903	TTTTTCCCTGACTCAGACAGCTATCTTATTTTTCAGAACTGTTTAAACCTTTGTGT	90844
QY	921	GCTGTTTAAATATATGTGTATATCTTGTGTGTTTCTTCCTGATATCAGAGCTGTTTCC	980
D	90843	GCTGTTTAAATATATGTGTATATCTTCTCATTTTCTGATATCAGAGCTGTTTCC	90784
QY	981	GTGGTGTGTAATATATTTTGTGTTTATATGTTTATATGGATGTTTATGATGTCAAGTT	1040





31766	31865	gap of unknown length
31866	32886	contig of 1021 bp in length
32887	32986	gap of unknown length
32987	34733	contig of 1747 bp in length
34734	34833	gap of unknown length
34834	35561	contig of 1728 bp in length
35562	36661	gap of unknown length
36662	38855	contig of 2194 bp in length
38856	38955	gap of unknown length
40676	40675	contig of 1720 bp in length
40776	40775	gap of unknown length
43314	43313	contig of 2538 bp in length
43414	43413	gap of unknown length
45971	45970	contig of 2557 bp in length
46071	46070	gap of unknown length
48911	48910	contig of 2840 bp in length
49011	49010	gap of unknown length
51896	51895	contig of 2885 bp in length
51996	51995	gap of unknown length
54124	54123	contig of 2128 bp in length
54224	54223	gap of unknown length
55764	55763	contig of 1540 bp in length
55864	55863	gap of unknown length
57513	57512	contig of 1649 bp in length
57613	57612	gap of unknown length
59523	59522	contig of 1910 bp in length
59623	59622	gap of unknown length
60938	60937	contig of 1315 bp in length
61038	61037	gap of unknown length
63207	63206	contig of 2169 bp in length
63307	63306	gap of unknown length
64956	64955	contig of 1650 bp in length
65057	65056	gap of unknown length
65057	67731	contig of 2675 bp in length
67732	67731	gap of unknown length
70618	70617	contig of 2786 bp in length
70718	70717	gap of unknown length
73448	73447	contig of 2730 bp in length
73548	73547	gap of unknown length
77163	77162	contig of 3615 bp in length
77263	77262	gap of unknown length
80488	80487	contig of 3225 bp in length
80588	80587	gap of unknown length
82853	82852	contig of 2265 bp in length
82953	82952	gap of unknown length
86852	86851	contig of 3899 bp in length
86952	86951	gap of unknown length
91800	91799	contig of 4848 bp in length
91900	91899	gap of unknown length
95212	95211	contig of 3312 bp in length
95312	95311	gap of unknown length
98113	98112	contig of 2801 bp in length
98213	98212	gap of unknown length
100746	100745	contig of 2533 bp in length
100846	100845	gap of unknown length
104056	104055	contig of 3210 bp in length
104156	104155	gap of unknown length
107670	107669	contig of 3514 bp in length
107770	107769	gap of unknown length
111041	111040	contig of 3271 bp in length
111141	111140	gap of unknown length
114379	114378	contig of 3238 bp in length
114479	114478	gap of unknown length
117845	117844	contig of 3367 bp in length

Query Match 40.8%; Score 573.8; DB 2; Length 174028;  
 Best Local Similarity 79.9%; Pred. No. 3,7e-83;  
 Matches 822; Conservative 0; Mismatches 97; Indels 110; Gaps 8;

OY 1 AAGGATACGAGT-GGGGGGCGCTGATGCTTATAGGAGCGAGTAGGGAGAGA 59  
 Db 100711 AAGGATACGAGTGGGGGCGCTGATGCTTATAGGAGCGAGTAGGGAGAGA 100653

OY 60	CGTTTCTCCCGGAGACGCTATCTCATCTCTTTCTTTGATTAACCGTGGCGGAG 119
Db 100652	CGTTTCTCTACTGGAACAGCTTATGCTATCCCTTCCATTAACCATGATGACGTA 100593
OY 120	AGTCAAGGCGGCGCTGGGAGACAGAGCGCGGTGGCGGCGGAGTGCAGTGC 179
Db 100592	GCTTGGGGCGGCGCTGGGAGAGAG--GGCGGTGGCGGAGGAGCGAGTGCAGTGC 100536
OY 180	ATGTCACGATGAATCCGGAATATGATTTATTAATCACTTCTGATGGGACATCA 239
Db 100535	ATGTCACGATGAATCCGGAAT--ATTATTTATTAATCACTTCTGATGGGATTC 100479
OY 240	GGGTTGGAAAGCTTGGCTTCTTATGATTTGATTCAGATGATATACGAAAGCTAC 299
Db 100478	GGGTTGGAAAGCTTGGCTTCTGCTTATGATTTGATTCAGATGATATACGAAAGCTAC 100419
OY 300	ATCAGCAATTTGGTGTGATTTCAAAATAGAACTATAGATTAAGAGGAAATATC 359
Db 100418	ATCAGCAATTTGGTGTGATTTCAAAATAGAACTATAGATTAAGAGGAAATATATC 100359
OY 360	AAGCTTCAATA----- 371
Db 100358	AAGCTTCAATA----- 371
OY 372	-----GAGTCTTCAAT 383
Db 100298	TACAGAGAACCATGCGCATATGATTTGATGATGATGACGAGACGAGACGCTTCAAT 100239
OY 384	AATGTTAAACAGTGGCTCAGAGAAATAGATGCTTATGCAATGAAATGTCAACAAATG 443
Db 100238	AATGTTAAACAGTGGCTCAGAGAAATAGATGCTTATGCAATGAAATGTCAACAAATG 100179
OY 444	TTGGTAGGAAACAATGTGATCTGCACCAAGAAAGTAGAGTACACAAACGAGAG 503
Db 100178	TTGGTAGGAAACAATGTGATCTGCACCAAGAAAGTAGAGTACACAAACGAGAG 100123
OY 504	GAATTTGCTGATTCCTTGAATTCGTTTGGAAACAGTGTAGAAATGCAACGAT 563
Db 100122	GAATTTGCTGATTCCTTGAATTCGTTTGGAAACAGTGTAGAAATGCAACGAT 100063
OY 564	GTAAGACAGTCTTATGATGATGAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 623
Db 100062	GTAAGACAGTCTTATGATGATGAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 100003
OY 624	ACAGTGTGCTGCTGAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 683
Db 100002	ACAGTGTGCTGCTGAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 99943
OY 684	GGTGAGAGTCTGCTTAATTTGCTCCATCCTTTCTCAGACCAATGAATTTGCAATC 743
Db 99942	GGTGAGAGTCTGCTTAATTTGCTCCATCCTTTCTCAGACCAATGAATTTGCAATC 99883
OY 744	TGAACCCAGTGAAGAAATTTGCTGATTTGATGATGATGATGATGATGATGATGAT 803
Db 99882	TGAACCCAGTGAAGAAATTTGCTGATTTGATGATGATGATGATGATGATGATGAT 99823
OY 804	GATTTCTACGCTGCTCAAAAGTGCAGAGTGAATGAATGAATGAATGAATGAATTT 862
Db 99822	GATTTCTACGCTGCTCAAAAGTGCAGAGTGAATGAATGAATGAATGAATGAATTT 99764
OY 863	TATTCCTTGAAGTCAAGACAGTCAATTTTCAAGTGTAACTTTTAACTTTGTGTGC 922
Db 99763	TATTCCTTGAAGTCAAGACAGTCAATTTTCAAGTGTAACTTTTAACTTTGTGTGC 99704
OY 923	TGTTTATA 931
Db 99703	TGTTTATA 99695

Search completed: November 17, 2002, 12:50:42  
 Job time : 4075 secs





KH Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
OS WO200175067-A2.  
PN 11-OCT-2001.  
PD 30-MAR-2001; 2001WO-US08631.  
PE 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
PI WPI: 2001-639362/73.  
DR P-PSDB: ABG19673.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
PS Claim 1: SEQ ID No 19664; 103pp: English.  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting and  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological functions.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on human  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX Sequence 1193 BP: 372 A; 279 C; 213 G; 329 T; 0 other;  
SQ  
Query Match 73.2%; Score 1029; DB 23; Length 1193;  
Best Local Similarity 89.9%; Pred. No. 8.7e-132;  
Matches 1182; Conservative 0; Mismatches 0; Indels 133; Gaps 2;

Db 943 |||||  
GGGCTTGAAAGCTCTTGGCTTCTTAGGTTGAGATGATACATATACGAAAGCTAC 884  
QY 300 ATCAGACAAATTGGTGTGGATTTCAAAATTAAGAACTATAGAGTTAGACGGGAAAAACATC 359  
Db 883 ATCAGACAAATTGGTGTGGATTTCAAAATTAAGAACTATAGAGTTAGACGGGAAAAACATC 824  
QY 360 AAGCTTCAAAATAGAGTCTTCCCTCAATTAATGTTAAACAGTGGCTGCAGAAATAGATGTTAT 419  
Db 823 AAGCTTCAAAATAGAGTCTTCCCTCAATTAATGTTAAACAGTGGCTGCAGAAATAGATGTTAT 812  
QY 420 GCCAGTGAATAATGTCACAAATTTGTTAGAGGAACAATGTGATCTGACCAAAAGAA 479  
Db 811 ----- 812  
QY 480 GTAGTACTACACACAGAGAAATTTGCTGATTCCTTGGAATTCGTTTTGGAA 539  
Db 811 -----GAAATTTGCTGATTCCTTGGAATTCGTTTTGGAA 776  
QY 540 ACCAGTCTTAAGAAATGCAAGAGAAATGCTTCTTCAATGAGATGAGAGCTGAGAT 599  
Db 775 ACCAGTCTTAAGAAATGCAAGAGAAATGCTTCTTCAATGAGATGAGAGCTGAGAT 716  
QY 600 AAAAAAGCAATGGGTCCCGGAGCAACAGCTGGTGTCTGAGAAAGTCAATGTTAAAT 659  
Db 715 AAAAAAGCAATGGGTCCCGGAGCAACAGCTGGTGTCTGAGAAAGTCAATGTTAAAT 656  
QY 660 CAGAGCACTCCAGTCAACAGCACTGAGAGGTGCTGTAATTTGGCTTCATCCTTT 719  
Db 655 CAGAGCACTCCAGTCAACAGCACTGAGAGGTGCTGTAATTTGGCTTCATCCTTT 596  
QY 720 TCTCAGCAATGAATTTGCAATCTGAAACCAAGTAAATTTGGCTGATTTGT 779  
Db 595 TCTCAGCAATGAATTTGCAATCTGAAACCAAGTAAATTTGGCTGATTTGT 536  
QY 780 ACTGTATAGTGTGACATACAGATTTCTACGCTCTCCAAAGAGTCAAGATTTGTA 839  
Db 535 ACTGTATAGTGTGACATACAGATTTCTACGCTCTCCAAAGAGTCAAGATTTGTA 476  
QY 840 ATGTGTCAATAGCAATTTTATTTTATTTCCCTTGACATCAAGAGTCAATTTGTA 899  
Db 475 ATGTGTCAATAGCAATTTTATTTTATTTCCCTTGACATCAAGAGTCAATTTGTA 416  
QY 900 AACTGTTTAAACCTTTGTGTGCTGGTTTAAATTAATGTGTATCTTGTGCTTT 959  
Db 415 AACTGTTTAAACCTTTGTGTGCTGGTTTAAATTAATGTGTATCTTGTGCTTT 356  
QY 960 CCTGATACCAAGCTTTCCCGGAGTGTGTTGTAATTAATTTGTTGATTTATAT 1019  
Db 355 CCTGATACCAAGCTTTCCCGGAGTGTGTTGTAATTAATTTGTTGATTTATAT 296  
QY 1020 GGCAATGTTAGATGCTGAGTTAGTCTGTAAGTGAAGTCAAGATTTGATTA 1079  
Db 295 GGCAATGTTAGATGCTGAGTTAGTCTGTAAGTGAAGTCAAGATTTGATTA 236  
QY 1080 CAGCAACAGCAGTGTCTGCTACCTTCCATGATTAAGTTAGTGAAGTTATAT 1139  
Db 235 CAGCAACAGCAGTGTCTGCTACCTTCCATGATTAAGTTAGTGAAGTTATAT 176  
QY 1140 ATCTGATTTGCTAGTTCTTCCCTGTAGAGTTAATTAATGAAGATTACATCTGATTA 1199  
Db 175 ATCTGATTTGCTAGTTCTTCCCTGTAGAGTTAATTAATGAAGATTACATCTGATTA 116  
QY 1200 ATATGTTCTTCACTACTGATTAATTTGCTGTCAGAAATTTGTAATTTGTTGACAA 1259  
Db 115 ATATGTTCTTCACTACTGATTAATTTGCTGTCAGAAATTTGTAATTTGTTGACAA 56  
QY 1260 CTATGTAAACAAACAACTGAAGATATGTTAATTAATTAATTTGTTGTTGAACT 1314  
Db 55 CTATGTAAACAAACAACTGAAGATATGTTAATTAATTAATTTGTTGTTGAACT 1

RESULT 3

ABK44823/C	ID	ABK44823 standard; cDNA; 506 BP.
ABK44823;	AC	
05-JUN-2002	DT	(first entry)
cdna encoding colon tumour protein, SEQ ID NO 374.	DE	
Human; colon tumour; vaccine; colon cancer; immunogenic;	KW	
immunotherapy; gene; ss.	XX	
Homo sapiens.	OS	
WO200212328-A2.	PN	
14-FEB-2002.	PD	
31-JUL-2001; 2001WO-US24218.	XX	
03-AUG-2000; 2000US-223283P.	PR	
28-MAR-2001; 2001US-279763P.	PR	
29-JUN-2001; 2001US-302051P.	PR	
(CORI-) CORIXA CORP.	PA	
Kling GE, Meagher MJ, Xu J, Secret H;	PI	
WPI; 2002-241739/29.	DR	
New colon cancer polypeptides and polynucleotides, useful as vaccines,	PT	
for diagnosing, preventing, and treating colon cancer, and as markers	PT	
for the progression of cancer -	PT	
Claim 1; SEQ ID NO 374; 147bp; English.	PS	
The invention relates to polynucleotides encoding colon tumour proteins.	XX	
The polynucleotides and encoded polypeptides are useful in pharmaceutical	CC	
compositions, such as vaccines, for the diagnosis, prevention, and	CC	
treatment of colon cancer. Polynucleotide sequences may be used as	CC	
hybridisation probes or primers, and in the design and preparation of	CC	
ribozyme molecules for inhibiting expression of tumour polypeptides and	CC	
proteins in tumour cells. The compositions are useful for stimulating an	CC	
immune response against cancer, particularly for the immunotherapy of	CC	
colon cancer, and as markers for the progression of cancer.	CC	
ABK44450-ABK46237 represent coding sequences of human colon tumour	CC	
proteins of the invention.	CC	
Note: With the exception of SEQ ID No 1 and 2, the sequence data	CC	
for this patent did not form part of the printed specification but was	CC	
supplied by the European Patent Office.	CC	
Sequence 506 BP; 194 A; 91 C; 82 G; 139 T; 0 other;	XX	

	Query Match	36.0%	Score 506	DB 24	Length 506
	Best Local Similarity	100.0%	Pred. NO. 1.2e-60		
	Matches 506	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	770	CCTGAATGTCCTGATGCTGACACACACATTTCTACCGTCTCCAAAGTCA	829		
Db	506	CCTGAATGTCCTGATGCTGACACACACATTTCTACCGTCTCCAAAGTCA	447		
QY	830	GAGATTGTAATGTCATACTGACTTTTTTTTATTCCTTGACTCAAGACGTA	889		
Db	446	GAGATTGTAATGTCATACTGACTTTTTTTTATTCCTTGACTCAAGACGTA	387		
QY	890	TCATTTTGAAGACGTTTTTAAACCTTGCGTCGTTATTAATATGCTGAATCC	949		
Db	386	TCATTTTGAAGACGTTTTTAAACCTTGCGTCGTTATTAATATGCTGAATCC	327		
QY	950	TTGTTGCTTTCCTGATACACAGACTTTTCCGCTGTTGGATATATTTGTTTGA	1009		
Db	326	TTGTTGCTTTCCTGATACACAGACTTTTCCGCTGTTGGATATATTTGTTTGA	267		

Qy	1010	TGTTATATTGGCATGTTTGGATGTACAGTTTAGTCTTGGAAGATGAAGTTCCAGCAATT	1069
Db	266	TGTTTATATTGGCATGTTTGGATGTACAGTTTAGTCTTGGAAGATGAAGTTCCAGCAATT	207
Qy	1070	TTGTATCAAAACAGCAGCAGCAGTGTCTGCATCTTCATGCATTAAGTTTAGTGAGAGT	1129
Db	206	TTGTATCAAAACAGCAGCAGCAGTGTCTGCATCTTCATGCATTAAGTTTAGTGAGAGT	147
Qy	1130	TATATGTAAGTCTGATTGCTGAGTCTTCCTGTGAGATTATAATGGAAAGATTACAC	1189
Db	146	TATATGTAAGTCTGATTGCTGAGTCTTCCTGTGAGATTATAATGGAAAGATTACAC	87
Qy	1190	TATCTGATTATAGTTCTTCATCTACTCTGCATTAATTGTGGCTGCAGAAATATTGTAA	1249
Db	86	TATCTGATTATAGTTCTTCATCTACTCTGCATTAATTGTGGCTGCAGAAATATTGTAA	27
Qy	1250	TTGTGGACACACTATGTAAACAAACAA 1275	
Db	26	TTGTGGACACACTATGTAAACAAACAA 1	

RESULT 4	
ID	ABK83978
XX	ABK83978 standard; cDNA: 723 bp.
XX	
AC	ABK83978;
XX	
DT	14-AUG-2002 (first entry)
XX	
DE	Human cDNA differentially expressed in granulocytic cells #549.
XX	
KW	Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW	viral infection; parasitic infection; protozoal infection;
KW	fungal infection; sterile inflammatory disease; psoriasis;
KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW	cardiac reperfusion injury; renal reperfusion injury; AIDS;
KW	adult respiratory distress syndrome; inflammatory bowel disease;
KW	Crohn's disease; ulcerative colitis; periodontal disease;
KW	granulocyte activation; chronic inflammation; allergy.
XX	
OS	Homo sapiens.
XX	
PN	WO200228999-A2.
XX	
PD	11-APR-2002.
XX	
PF	03-OCT-2001; 2001WO-US30821.
XX	
PR	03-OCT-2000; 2000US-237189P.
XX	
PA	(GENE-) GENE LOGIC INC.
XX	
PI	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX	
DR	WPI; 2002-435328/46.
XX	
PT	Detecting granulocyte activation by detecting differential expression
PT	of genes associated with granulocyte activation, which serves as
PT	diagnostic markers that is useful for monitoring disease states and
PT	drug toxicity
XX	
PS	Claim 1; SEQ ID No 549; 114pp; English.
XX	
XX	The invention relates to detecting (M1) granulocyte (GC) activation
CC	(GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC	DNA chip analysis as given in the specification, and comparing
CC	the expression level to an expression level in an unactivated
CC	GC, where differential expression of Gs is indicative of GCA.
CC	Also included are modulating (M2) GA by contacting GC with an agent
CC	that alters the expression of at least one gene in Gs; (2) screening (M3)
CC	for an agent capable of modulating GCA or an inflammation (especially
CC	chronic) in a tissue, an allergic response in a subject, exposure of a
CC	subject to a pathogen or sterile inflammatory disease using the



Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.2%; Score 467; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 2.4e-55;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGCTGCTAAGATTCACAA 561  
 DB 467 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGCTGCTAAGATTCACAA 408  
 QY 562 ATGAGAACAGCTTTTCATGACGATGAGATGATTAATAAACGATGGGTCGCGAG 621  
 DB 407 ATGAGAACAGCTTTTCATGACGATGAGATGATTAATAAACGATGGGTCGCGAG 348  
 QY 622 CAACAGCTGGTGTGCTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGAGT 681  
 DB 347 CAACAGCTGGTGTGCTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGAGT 288  
 QY 682 CAGGTGAGGTGCTGCTAAATTTGCTCCTCATCTTTTTCACAGCAATGATTTGCAA 741  
 DB 287 CAGGTGAGGTGCTGCTAAATTTGCTCCTCATCTTTTTCACAGCAATGATTTGCAA 228  
 QY 742 TCTGAACCCCAAGTGAATAATGCTGAAATGATGATGATGATGATGATGATGATGAT 801  
 DB 227 TCTGAACCCCAAGTGAATAATGCTGAAATGATGATGATGATGATGATGATGATGAT 168  
 QY 802 CAGATTTCTACCGTCTCCACAAAGTGCAGAGATTTGAATGATGATGATGATGATGAT 861  
 DB 167 CAGATTTCTACCGTCTCCACAAAGTGCAGAGATTTGAATGATGATGATGATGATGAT 108  
 QY 862 TTATTCCTTGACCTCAAGACAGCTAATCTATTTTCAGAACTGTTTAACTTTGCTG 921  
 DB 107 TTATTCCTTGACCTCAAGACAGCTAATCTATTTTCAGAACTGTTTAACTTTGCTG 48  
 QY 922 CTGGTTTAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968  
 DB 47 CTGGTTTAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1

RESULT 6  
 ABA53512/c  
 ID ABA53512 standard; DNA: 487 BP.

AC ABA53512;

DT 01-FEB-2002 (first entry)

DE Human foetal liver single exon nucleic acid probe #1817.

KW Human: foetal liver: gene expression; single exon nucleic acid probe; ss.

OS Homo sapiens.

PN WO200157277-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001MO-US00669.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483447/52.

PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX Claim 1; SEQ ID NO 1817; 639pp + sequence listing; English.

CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC foetal liver. The present sequence is a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.2%; Score 467; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 2.4e-55;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGCTGCTAAGATTCACAA 561  
 DB 467 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGCTGCTAAGATTCACAA 408  
 QY 562 ATGAGAACAGCTTTTCATGACGATGAGATGATTAATAAACGATGGGTCGCGAG 621  
 DB 407 ATGAGAACAGCTTTTCATGACGATGAGATGATTAATAAACGATGGGTCGCGAG 348  
 QY 622 CAACAGCTGGTGTGCTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGAGT 681  
 DB 347 CAACAGCTGGTGTGCTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGAGT 288  
 QY 682 CAGGTGAGGTGCTGCTAAATTTGCTCCTCATCTTTTTCACAGCAATGATTTGCAA 741  
 DB 287 CAGGTGAGGTGCTGCTAAATTTGCTCCTCATCTTTTTCACAGCAATGATTTGCAA 228  
 QY 742 TCTGAACCCCAAGTGAATAATGCTGAAATGATGATGATGATGATGATGATGATGAT 801  
 DB 227 TCTGAACCCCAAGTGAATAATGCTGAAATGATGATGATGATGATGATGATGATGAT 168  
 QY 802 CAGATTTCTACCGTCTCCACAAAGTGCAGAGATTTGAATGATGATGATGATGATGAT 861  
 DB 167 CAGATTTCTACCGTCTCCACAAAGTGCAGAGATTTGAATGATGATGATGATGATGAT 108  
 QY 862 TTATTCCTTGACCTCAAGACAGCTAATCTATTTTCAGAACTGTTTAACTTTGCTG 921  
 DB 107 TTATTCCTTGACCTCAAGACAGCTAATCTATTTTCAGAACTGTTTAACTTTGCTG 48  
 QY 922 CTGGTTTAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968  
 DB 47 CTGGTTTAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1

RESULT 7  
 ABA23276/c  
 ID ABA23276 standard; DNA: 487 BP.

AC ABA23276;

DT 23-JAN-2002 (first entry)

DE Probe #1742 for gene expression analysis in human heart cell sample.

KW Human: gene expression; heart; microarray; vascular system; probe;

KW cardiovascular disease; hypertension; cardiac arrhythmia;

KW congenital heart disease; ss.

OS Homo sapiens.

PN WO200157274-A2.

PD 09-AUG-2001.



```
XX 30-JAN-2001; 2001WO-US00666.
PR
XX 04-FEB-2000; 2000US-0180312.
FR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 1: SEQ ID NO 1742; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart. The
XX present sequence is one such probe. The probes may be used for
XX predicting, measuring and displaying gene expression in samples derived
XX from the human heart via microarrays. By measuring gene expression, the
XX probes are useful for predicting, diagnosing, grading, staging,
XX monitoring and prognosing diseases of the human heart and vascular system
XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
XX congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match      33.2%; Score 467; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 2.4e-55;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGTCTTAAGATGCAACGA 561
   |||||||
DB 467 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGTCTTAAGATGCAACGA 408
QY 562 ATGTGAACAGCTTTTCATGACGATGACGATGATTAATAAAGCAATGGTCCCGAG 621
   |||||||
DB 407 ATGTGAACAGCTTTTCATGACGATGACGATGATTAATAAAGCAATGGTCCCGAG 348
QY 622 CAACAGCTGTGTGCTGAGAGTCGCAATGTTAAATTCAGACACTCCAGTCAAGCAGT 681
   |||||||
DB 347 CAACAGCTGTGTGCTGAGAGTCGCAATGTTAAATTCAGACACTCCAGTCAAGCAGT 288
QY 682 CAGGTGAGGTTGCTCTTAATAATTTGCTTCAATCTTTTCTACAGCAATGATTTGCA 741
   |||||||
DB 287 CAGGTGAGGTTGCTCTTAATAATTTGCTTCAATCTTTTCTACAGCAATGATTTGCA 228
QY 742 TCTGAACCAAGTGAAAAACAATAATGCTGATGATGATGATGATGATGATGATGATGAT 801
   |||||||
DB 227 TCTGAACCAAGTGAAAAACAATAATGCTGATGATGATGATGATGATGATGATGATGAT 168
QY 802 CAGATTTTACCTCTCCACAAAGGTGAGATGATGATGATGATGATGATGATGATGATGAT 861
   |||||||
DB 167 CAGATTTTACCTCTCCACAAAGGTGAGATGATGATGATGATGATGATGATGATGATGAT 108
QY 862 TTAATTCCTTGACTCAAGAGCTAATCTATTTTCAAACTGTTTAAACCTTTGTGTG 921
   |||||||
DB 107 TTAATTCCTTGACTCAAGAGCTAATCTATTTTCAAACTGTTTAAACCTTTGTGTG 48
QY 922 CTGGTTTAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 968
   |||||||
DB 47 CTGGTTTAAATAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1
```

```
RESULT 8
AAK01783/C
ID AAK01783 standard; DNA; 487 BP.
XX
XX AAK01783:
AC
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe SEQ ID NO: 1774.
XX
XX Human brain expressed single exon probe SEQ ID NO: 1774.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -
XX
XX Example 4; SEQ ID NO: 1774; 650pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is one of the probes of the
XX invention.
XX
SQ Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match      33.2%; Score 467; DB 22; Length 487;
Best Local Similarity 100.0%; Pred. No. 2.4e-55;
Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGTCTTAAGATGCAACGA 561
   |||||||
DB 467 AGGAATTTGCTGATTCCTTGGAAATTCCTTTTGGAAACCAAGTCTTAAGATGCAACGA 408
QY 562 ATGTGAACAGCTTTTCATGACGATGACGATGATTAATAAAGCAATGGTCCCGAG 621
   |||||||
DB 407 ATGTGAACAGCTTTTCATGACGATGACGATGATTAATAAAGCAATGGTCCCGAG 348
QY 622 CAACAGCTGTGTGCTGAGAGTCGCAATGTTAAATTCAGACACTCCAGTCAAGCAGT 681
   |||||||
DB 347 CAACAGCTGTGTGCTGAGAGTCGCAATGTTAAATTCAGACACTCCAGTCAAGCAGT 288
QY 682 CAGGTGAGGTTGCTCTTAATAATTTGCTTCAATCTTTTCTACAGCAATGATTTGCA 741
   |||||||
DB 287 CAGGTGAGGTTGCTCTTAATAATTTGCTTCAATCTTTTCTACAGCAATGATTTGCA 228
QY 742 TCTGAACCAAGTGAAAAACAATAATGCTGATGATGATGATGATGATGATGATGATGATGAT 801
```





PN W0200157270-A2.  
 XX 09-AUG-2001.  
 XX 29-JAN-2001; 2001WO-US00661.  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 XX  
 PS Claim 25; SEQ ID No 1743; 322pp; English.  
 CC The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridises at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,  
 CC staging, monitoring and prognosing diseases of the human breast,  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocytic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 CC  
 XX Sequence 487 BP; 154 A; 97 C; 98 G; 138 T; 0 other;

Query Match 33.2%; Score 467; DB 22; Length 487;

Best Local Similarity 100.0%; Pred. No. 2,4e-55;

Matches 467; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 502 AGGAATTTGCGATTCCTTGGAAATCCGTTTGGAAACGAGTGAAGTGAACGA 561  
 |||||||  
 DB 467 AGGAATTTGCGATTCCTTGGAAATCCGTTTGGAAACGAGTGAAGTGAACGA 408  
 |||||||  
 QY 562 ATGTAGAACACTCTTTCATGACGATGCGAGTGAATTAAGCGAATGGTCCGGAG 621  
 |||||||  
 DB 407 ATGTAGAACACTCTTTCATGACGATGCGAGTGAATTAAGCGAATGGTCCGGAG 348  
 |||||||  
 QY 622 CAACAGCGTGGTGGTGCAGAGTCCATGTAAATTCACAGACCTCCAGTCAAGCAGT 681  
 |||||||  
 DB 347 CAACAGCGTGGTGGTGCAGAGTCCATGTAAATTCACAGACCTCCAGTCAAGCAGT 288  
 |||||||  
 QY 682 CAGGTGAGTGGTGGTGAATTTGGCTCCATCTTTTTCACAGCATGAATTTGCAA 741  
 |||||||  
 DB 287 CAGGTGAGTGGTGGTGAATTTGGCTCCATCTTTTTCACAGCATGAATTTGCAA 228  
 |||||||  
 QY 742 TCTGAACCAAGTGAAGTCAAGTGTATGTATGTAGTGCACACTACAA 801  
 |||||||  
 DB 227 TCTGAACCAAGTGAAGTCAAGTGTATGTATGTAGTGCACACTACAA 168  
 |||||||  
 QY 802 CAGATTTTACCGTCCACCAAGTCAAGTGTATGTATGTATGTATGTATGTAT 861  
 |||||||  
 DB 167 CAGATTTTACCGTCCACCAAGTCAAGTGTATGTATGTATGTATGTATGTAT 108  
 |||||||  
 QY 862 TTATTTCCCTTACACAGACGACTAATTTTTCAGACGTTTAACTTTTGTGTG 921  
 |||||||  
 DB 107 TTATTTCCCTTACACAGACGACTAATTTTTCAGACGTTTAACTTTTGTGTG 48

QY 922 CTGTTTATTAATAATGTGTATTCCTTGTCTTCTGATACC 968  
 |||||||  
 DB 47 CTGTTTATTAATAATGTGTATTCCTTGTCTTCTGATACC 1  
 |||||||  
 RESULT 13  
 ID ABS01772/c  
 ID ABS01772 standard; DNA; 487 BP.  
 AC ABS01772;  
 XX 19-AUG-2002 (first entry)  
 XX  
 XX Human genome-derived single exon probe from lung SEQ ID No 1763.  
 DE Human; ds; single exon probe; asthma; lung cancer; COPD; IHD;  
 KW Chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 KW  
 XX Homo sapiens.  
 OS  
 XX W0200186003-A2.  
 PN 15-NOV-2001.  
 PD  
 XX 30-JAN-2001; 2001WO-US00665.  
 XX  
 XX 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PR  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 PT  
 PS Claim 1; SEQ ID No 1763; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of



DT 13-SEP-2002 (first entry)  
XX  
DE Human prostate expression marker CDNA 14052.  
XX  
KW Human: prostate cancer: cytostatic; carcinogen; pharmacodynamic marker;  
KW pharmacogenomic marker; gene; ss.  
XX  
OS Homo sapiens.  
PN WO200160860-A2.  
XX  
PD 23-AUG-2001.  
XX  
PF 20-FEB-2001; 2001WO-US05171.  
XX  
PR 17-FEB-2000; 2000US-183319P.  
PR 16-MAR-2000; 2000US-189862P.  
PR 23-MAY-2000; 2000US-207454P.  
PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Endege WO, Monahan JE;  
XX  
DR WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer -  
XX  
PS Claim 1; Page 2343-2344; 11750pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule (I) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (I) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 406 BP; 104 A; 99 C; 92 G; 111 T; 0 other;  
XX  
Query Match 18.8%; Score 264; DB 23; Length 406;  
Best Local Similarity 100.0%; Pred. No. 9.7e-28;  
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 518 CCTGGAATTCGTTTGGAAACAGTCTAGAAATGCAACGAATGTAGAACAGTCTT 577  
DB 406 CCTTGAATTCGTTTGGAAACAGTCTAGAAATGCAACGAATGTAGAACAGTCTT 347  
QY 578 CATGACATGCGAGCTGAGATTAAAGCAATGGGTCGAGACACAGCTGTGTGTC 637  
DB 346 CATGACATGCGAGCTGAGATTAAAGCAATGGGTCGAGACACAGCTGTGTGTC 287  
QY 638 TGAGAGTCCATGTTAAATTCAGAGCAGTCAGTCAAGCAGTGAAGTGTGCTG 697  
DB 286 TGAGAGTCCATGTTAAATTCAGAGCAGTCAGTCAAGCAGTGAAGTGTGCTG 227  
QY 698 CTAATAATTTGCTCATCTTTTCCACAGCAATGAATTCGAATCTGAACCAAGTGAA 757  
DB 226 CTAATAATTTGCTCATCTTTTCCACAGCAATGAATTCGAATCTGAACCAAGTGAA 167  
QY 758 AAAACAAAATTCCTGAATGTAC 781

DB 166 AAAACAAAATTCCTGAATGTAC 143

Search completed: November 17, 2002, 11:42:33  
Job time : 261 secs



D 241 GGGTTGGAAAGCTTGCCTCTCTAGTTGAGATGATATACAGAAAGCTACA 300  
Q 301 TCACACAAATGTGTGGATTTCAAAATTAAGACTATAGAGTACGCGGAAACATCA 360  
D 301 TCACACAAATGTGTGGATTTCAAAATTAAGACTATAGAGTACGCGGAAACATCA 360  
Q 361 AGCTTCAAAATAGAGTCTTCAATATATTAACAGTGGCTGAGAAATAGATGTTATG 420  
D 361 AGCTTCAAAATAGAGTCTTCAATATATTAACAGTGGCTGAGAAATAGATGTTATG 420  
Q 421 CCAGTGAATAATGTCAACAAATTTGTGTAGGAGCAAAATGTGATCTGACCAAGAAAG 480  
D 421 CCAGTGAATAATGTCAACAAATTTGTGTAGGAGCAAAATGTGATCTGACCAAGAAAG 480  
Q 481 TAGTAGACTACACAACAGCGAAGAAATTTGCTGATTCCTTGGAAATCCGTTTGGAAA 540  
D 481 TAGTAGACTACACAACAGCGAAGAAATTTGCTGATTCCTTGGAAATCCGTTTGGAAA 540  
Q 541 CCAATGCTAGAAATGCAACGAATGTAGAACAGTCTTTCATGACGATGCGAGCTGAGATTA 600  
D 541 CCAATGCTAGAAATGCAACGAATGTAGAACAGTCTTTCATGACGATGCGAGCTGAGATTA 600  
Q 601 AAAAGCGAATGGGTCCCGAGACACAGCTGGTGTGCTGAGAAATCCAAATGTTAAATTC 660  
D 601 AAAAGCGAATGGGTCCCGAGACACAGCTGGTGTGCTGAGAAATCCAAATGTTAAATTC 660  
Q 661 AGAGCACTCCAGTCAAGCAGTCAAGTGGAGGTGCTGCTAAATTTGCTCATCTCTTT 720  
D 661 AGAGCACTCCAGTCAAGCAGTCAAGTGGAGGTGCTGCTAAATTTGCTCATCTCTTT 720  
Q 721 CTCACAGCAATGAAATTTGCAATCTGAAACCAAGTGAAGAAACAAATTTGCTCATCTTT 780  
D 721 CTCACAGCAATGAAATTTGCAATCTGAAACCAAGTGAAGAAACAAATTTGCTCATCTTT 780  
Q 781 CTGATAGTAGTGGACATCAACAGATCTTACCGCTCCCAAAAGGTGAGATGTTAA 840  
D 781 CTGATAGTAGTGGACATCAACAGATCTTACCGCTCCCAAAAGGTGAGATGTTAA 840  
Q 841 TGGTCAATCTGACTTTTATTTTATTCCTTGACTCAAGACAGTAACTTATTTTGA 900  
D 841 TGGTCAATCTGACTTTTATTTTATTCCTTGACTCAAGACAGTAACTTATTTTGA 900  
Q 901 ACTGTTTAAACCTTTGTGTGCTGTTTAAATATGTGTATTCCTTGTGCTTTC 960  
D 901 ACTGTTTAAACCTTTGTGTGCTGTTTAAATATGTGTATTCCTTGTGCTTTC 960  
Q 961 CTGATACAGCTGTTCCTGCTGGTGTGATGATATTTTGTGATGATTTATATG 1020  
D 961 CTGATACAGCTGTTCCTGCTGGTGTGATGATATTTTGTGATGATTTATATG 1020  
Q 1021 GCATGTTAGATGTCAGGTTAGTCTTCTGAAGATGAAGTTCAGCCATTTGTATCAAC 1080  
D 1021 GCATGTTAGATGTCAGGTTAGTCTTCTGAAGATGAAGTTCAGCCATTTGTATCAAC 1080  
Q 1081 AGCAACAGAGCTGTCTGCTCACTTTCATGATTAAGTTTATGAGATGTATATGTAAGA 1140  
D 1081 AGCAACAGAGCTGTCTGCTCACTTTCATGATTAAGTTTATGAGATGTATATGTAAGA 1140  
Q 1141 TCTGATTTGCTAGTCTTCTCTGTAGAGTTTAAATGGAAGATTTACCTATCGATTAA 1200  
D 1141 TCTGATTTGCTAGTCTTCTCTGTGTAGAGTTTAAATGGAAGATTTACCTATCGATTAA 1200  
Q 1201 TAGTTTCTCACTGCTGATATATTTGTGCTGAGAAATTTGTAATTTGTTGCAAC 1260  
D 1201 TAGTTTCTCACTGCTGATATATTTGTGCTGAGAAATTTGTAATTTGTTGCAAC 1260  
Q 1261 TATGTAAACAAACAGTGAAGATATGTTTAAATATTTAGCTTATTTGGAAGTAA 1320  
D 1261 TATGTAAACAAACAGTGAAGATATGTTTAAATATTTAGCTTATTTGGAAGTAA 1320  
Q 1321 AA 1380  
D 1321 AA 1380

Q 1381 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1405  
D 1381 AAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1405

RESULT 2  
US-09-820-003a-3  
: Sequence 3, Application US/09820003A  
: Patent No. US20020142382A1  
: GENERAL INFORMATION:  
: APPLICANT: MERKULOV, Gennady et al.  
: TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
: TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
: FILE REFERENCE: CL001196  
: CURRENT APPLICATION NUMBER: US/09/820,003A  
: CURRENT FILING DATE: 2001-03-29  
: NUMBER OF SEQ ID NOS: 39  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 3  
: LENGTH: 46050  
: TYPE: DNA  
: ORGANISM: Homo sapien  
: FEATURE:  
: NAME/KEY: misc\_feature  
: LOCATION: (1)..(46050)  
: OTHER INFORMATION: n = A,T,C or G  
US-09-820-003a-3

Query Match 58.28; Score 817.2; DB 10; Length 46050;  
Best Local Similarity 99.6%; Pred. No. 2,7e-105;  
Matches 819; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Q 502 AGAATTTGCTGATTCCTTCCGTAATTCCTTTTGGAAACAGGCTAAAGATGCAACGA 561  
D 43238 AGAATTTGCTGATTCCTTCCGTAATTCCTTTTGGAAACAGGCTAAAGATGCAACGA 43297  
Q 562 ATGTAGAACAATCTTTCATGACGATGACGAGTGAATTAAGAGGAGTCCGAG 621  
D 43298 ATGTAGAACAATCTTTCATGACGATGACGAGTGAATTAAGAGGAGTCCGAG 43357  
Q 622 CAACAGCTGTGTGTGCTGAGAGTCCAAATGTTAAATTCAGAGCAGTCCAGCAGT 681  
D 43358 CAACAGCTGTGTGTGCTGAGAGTCCAAATGTTAAATTCAGAGCAGTCCAGCAGT 43417  
Q 682 CAGGTGAGAGTGTGCTGTAATTTGCTCCATCCCTTTTCCAGAGAAATGTAATGCAA 741  
D 43418 CAGGTGAGAGTGTGCTGTAATTTGCTCCATCCCTTTTCCAGAGAAATGTAATGCAA 43477  
Q 742 TCTGAACCAAGTGAACAAATTCCTGATTTGATGATGATGCTGACTACAA 801  
D 43478 TCTGAACCAAGTGAACAAATTCCTGATTTGATGATGATGCTGACTACAA 43537  
Q 802 CAGATTTCTACGCTCTCCACAAAGGTGAGATGTAATGCTCAATCTGACTTTT 861  
D 43538 CAGATTTCTACGCTCTCCACAAAGGTGAGATGTAATGCTCAATCTGACTTTT 43597  
Q 862 TTATTTCCCTGACCAAGCAGTAACTTCACTTTTCAAGACTGTTTAACTTTGCTG 921  
D 43598 TTATTTCCCTGACCAAGCAGTAACTTCACTTTTCAAGACTGTTTAACTTTGCTG 43657  
Q 922 CTGTTTATAAATAATGTGTGTAATCTTGTCTTCCGATGACAGACTGTTCCG 981  
D 43658 CTGTTTATAAATAATGTGTGTAATCTTGTCTTCCGATGACAGACTGTTCCG 43717  
Q 982 TGGTTGTTAGAAATATTTTGTGATGTTTAAATGCGATGTTAGATGTCAGGTT 1041  
D 43718 TGGTTGTTAGAAATATTTTGTGATGTTTAAATGCGATGTTAGATGTCAGGTT 43777  
Q 1042 AGCTTTCTGAAGATGAAGTTCAGCATTGTTATCAACACAGCAGCAGTGTGTCAC 1101  
D 43778 AGCTTTCTGAAGATGAAGTTCAGCATTGTTATCAACACAGCAGCAGTGTGTCAC 43837





```

RESULT 5
US-10-033-528-374/C
: Sequence 374, Application US/10033528
: Patent No. US200201971A1
: GENERAL INFORMATION:
: APPLICANT: King, Gordon E.
: APPLICANT: Meagher, Madeline Joy
: APPLICANT: Xu, Jiangchun
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY.
: TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
: FILE REFERENCE: 210121.54/CI
: CURRENT APPLICATION NUMBER: US/10/033.528
: CURRENT FILING DATE: 2001-12-26
: NUMBER OF SEQ ID NOS: 1896
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 374
: LENGTH: 506
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-033-528-374

```

```

Db 347 CAACAGCTGGTGGTGAAGTCAATGTAATTCAGACGACTCCAGCAAGCAAGT 288
QY 682 CAGGTGAGGTTGCTGCTGTAATTTGCTTCACATCTTTTCTCAGACAAATGAAATTCGA 741
Db 287 CAGGTGAGGTTGCTGCTGTAATTTGCTTCACATCTTTTCTCAGACAAATGAAATTCGA 228
QY 742 TCTGAACCCAGTGAAGAAACAAATTTGCTGTAATTTGCTTCAGACAAATGAAATTCGA 801
Db 227 TCTGAACCCAGTGAAGAAACAAATTTGCTGTAATTTGCTTCAGACAAATGAAATTCGA 168
QY 802 CAGATCTTACGCTCCCAACAAAGCTAGAGATTTGTAATTTGCTTCAGACAAATGAAATTCGA 861
Db 167 CAGATCTTACGCTCCCAACAAAGCTAGAGATTTGTAATTTGCTTCAGACAAATGAAATTCGA 108
QY 862 TTAATTCCTTGAAGTCAAGACGACTACATTTTCAAGATGTTTAAACCTTTTGTG 921
Db 107 TTAATTCCTTGAAGTCAAGACGACTACATTTTCAAGATGTTTAAACCTTTTGTG 48
QY 922 CTGATTTAATAATATGTTGTAATCTGTTGCTTCTGATAC 968
Db 47 CTGATTTAATAATATGTTGTAATCTGTTGCTTCTGATAC 1

RESULT 7
US-09-960-352-5427
; Sequence 5427, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5427
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 23-LIB34-061-Q1-E1-F3
US-09-960-352-5427

Query Match 24.0%; Score 337.2; DB 10; Length 353;
Best Local Similarity 98.9%; Pred. No. 4,8e-39;
Matches 350; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 946 ATCCTTGTCTTCTGCTGATACAGACTGTTCCCGTGGTGGTGAATATATTTGTT 1005
Db 1 ATCCTTGTCTTCTGCTGATACAGACTGTTCC-GTGGTGGTGAATATATTTGTT 59
QY 1006 TTGATGTTTATATGTCATGTTAGATGTCAGTTTACTTCTGAGATGGAAGTTCAGC 1065
Db 60 TTGATGTTTATATGTCATGTTAGATGTCAGTTTACTTCTGAGATGGAAGTTCAGC 119
QY 1066 CATTTTATCAACAGACAGACAGAGCTGTCTCACTTCCATGACATGAAGTTAGAG 1125
Db 120 CATTTTATCAACAGACAGACAGAGCTGTCTCACTTCCATGACATGAAGTTAGAG 179
QY 1126 ATGTTAATATAGATGATTTCTGATTTCTCTTCTGAGATGTTAAATGAAAGTT 1185
Db 180 ATGTTAATATAGATGATTTCTGATTTCTCTTCTGAGATGTTAAATGAAAGTT 229
QY 1186 ACAGATATCATTAATAGTTTCTCATACATGTCATATAATTTGGGCGCAAGATTTG 1245
Db 240 ACAGATATCATTAATAGTTTCTCATACATGTCATATAATTTGGGCGCAAGATTTG 299
QY 1246 TAAATTTGTCACACATGATTAACAAACAGTGAAGATGTTAATTAATTT 1299
Db 300 TAAATTTGTCACACATGATTAACAAACAGTGAAGATGTTAATTAATTT 353

```

```

RESULT 8
US-09-960-352-8911
; Sequence 8911, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8911
; LENGTH: 427
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 38-LIB34-021-Q1-E1-B6
US-09-960-352-8911

Query Match 22.4%; Score 314.8; DB 10; Length 427;
Best Local Similarity 94.7%; Pred. No. 6e-36;
Matches 337; Conservative 0; Mismatches 17; Indels 2; Gaps 1;

QY 18 GCGGCTGCTGATTTGTTCTAGGGAGCGAGTACGGG--AAGACGTTTCTCTCCGGA 75
Db 1 GCGGCTGCTGATTTGTTCTAGGGAGCGAGTACGGGAGTGGGGTAAAGAGTTTCTCTGGA 60
QY 76 ACAGCTATATCATTTCTTCTTCTGATTTACCGTGGCGGAGACATGACAGGCGGCT 135
Db 61 ACAGCTATATCATTTCTTCTTCTGATTTACCGTGGCGGAGTGGGGTAAAGAGTTTCTCTG 120
QY 136 GCGGACGACAGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 195
Db 121 GCGGACGACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
QY 196 CGCATATGATTTATTTATGATTTACTTCTGATTTGCGACTCAGAGGTTGGAAAGTCT 255
Db 181 CGCATATGATTTATTTATGATTTACTTCTGATTTGCGACTCAGAGGTTGGAAAGTCT 240
QY 256 GCGCTTCTTCTGATTTGTCATGATTTACATATACAGAAAGCTACATGACAAATTTG 315
Db 241 GCGCTTCTTCTGATTTGTCATGATTTACATATACAGAAAGCTACATGACAAATTTG 300
QY 316 TGGATTTCAAAATTAAGAACTATAGATTAGACGAGGAAACAAATCAAGCTTCAATA 371
Db 301 TGGATTTCAAAATTAAGAACTATAGATTAGGTGGAACAAATCAAGCTTCAAGATA 356

RESULT 9
US-09-294-093B-4584
; Sequence 4584, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laludl, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4584
; LENGTH: 290
; TYPE: DNA
; ORGANISM: Zea mays

```





```

; APPLICANT: Zehentner, Barbara
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.491C7
; CURRENT APPLICATION NUMBER: US/10/010.742
; CURRENT FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 307
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 163, 191, 203, 222, 223, 228
; OTHER INFORMATION: n = A,T,C or G
US-10-010-742-84

```

```

Query Match      15.6%; Score 219.8; DB 12; Length 229;
Best Local Similarity 96.5%; Pred. No. 7.9e-23;
Matches 221; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 780 ACTGTATGATGCTGCTACACAGATTTTACCGTCTCCACAAAGTTCAGAGATTGTAA 839
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 ACTGTTGTAGCTGCTACACAGATTTTACCGTCTCCACAAAGTTCAGAGATTGTAA 60

QY 840 ATGTCATATCTGACTTTTATTTATCCCTGCTCAGACAGCTAATCTTCAATTTTCAG 899
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ATGTCATATCTGACTTTTATTTATCCCTGCTCAGACAGCTAATCTTCAATTTTCAG 120

QY 900 AACTGTTTAAACCTTTGTGCTGCTTTTAAATAATGCTTAATCCTTTGCTTT 959
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 AACTGTTTAAACCTTTGTGCTGCTTTTAAATAATGCTTAATCCTTTGCTTT 180

QY 960 CCTGATACACAGCTGTTCCCGTGGTGGTTAGAAATATATTTTGTGTTTG 1008
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 CCTGATACACAGCTGTTCCCGTGGTGGTTAGAAATATATTTTGTGTTTG 229

```

Search completed: November 17, 2002, 13:04:08  
 Job time : 132 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2002, 11:36:02 ; Search time 1933 Seconds  
(without alignments)  
11771.681 Million cell updates/sec

Title: US-09-820-003a-1

Perfect score: 1405  
1 aagcagatagcgtgctgcgc.....aaaaaaaaaaaaaaaa 1405

Scoring table: IDENTITY\_MNC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_mam:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	743	52.9	948	9	AL539022
2	732.4	52.1	777	14	B0014597
3	727.6	51.8	811	9	AA746643
4	723.8	51.5	817	13	BI869965
5	708.8	50.4	910	14	BQ276678
6	708.4	50.4	761	14	BQ772048

7	707.8	50.4	945	9	AL530265
8	704.4	50.1	1105	13	BM452262
9	703.6	50.1	725	14	BQ448090
10	702	50.0	796	13	BI913092
11	692.8	49.3	760	14	BM971301
12	691.8	49.2	976	13	BM919860
13	690.6	49.2	837	13	BI086445
14	689	49.0	689	14	BM706159
15	688.2	49.0	737	13	BI092003
16	686	48.8	686	14	BM826571
17	685.6	48.8	717	9	AI952757
18	683.2	48.6	706	14	BM975245
19	682.6	48.6	796	13	BI659520
20	677.6	48.2	877	12	BG617277
21	676.4	48.1	678	13	BG393358
22	671.2	47.8	684	13	BI495589
23	670.8	47.7	1130	13	BM542833
24	669.6	47.7	813	14	BQ178972
25	668.2	47.6	812	14	BQ178972
26	661.2	47.1	666	14	BM599300
27	659.6	46.9	721	9	AI884351
28	658	46.8	733	12	BQ178258
29	657.6	46.8	680	10	AM150848
30	655.2	46.5	682	14	BQ010358
31	650.4	46.3	692	9	AI829874
32	650	46.3	908	9	AA559410
33	648	46.1	723	9	AI800097
34	645.6	46.0	682	9	AI858694
35	640.4	45.5	900	12	BQ032817
36	638.8	45.5	682	9	AI860577
37	638	45.4	638	14	BM783973
38	638	45.4	646	14	BQ599347
39	638	45.4	689	14	BE646347
40	637.4	45.4	719	14	BQ210709
41	632.8	45.0	1055	13	BM472577
42	631.8	44.9	664	10	BE617921
43	631.2	44.9	664	10	BQ232373
44	630.6	44.9	645	14	AM166902
45	628.6	44.7	702	9	AI889183

# ALIGNMENTS

RESULT 1  
AL539022  
LOCUS AL539022  
DEFINITION AL539022 LTR FL013.FBRI Homo sapiens CDNA clone CSDF030YG06 5  
ACCESSION AL539022.1 GI:12867866  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 948)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
source  
1. 948  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CSDF030YG06"  
/clone\_lib="LTR\_FL013\_FBRI"  
/dev\_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"  
/lab\_host="DH10B"





QY 613 GTCGGGAGACAGCTGGTGGCTGAGAGATCAATTAATTAATTCAGACACTCCAG 672  
 DB 718 GTCCGGGAGACAGCTGGTGGCTGAGAGATCAATTAATTAATTCAGACACTCCAG 659  
 QY 673 TCAGAGAGTCAAGTGGAGTGGCTGCTAATAATTTGGCTCCATCTTTTCTCAGACAGATG 732  
 DB 658 TCAGAGAGTCAAGTGGAGTGGCTGCTAATAATTTGGCTCCATCTTTTCTCAGACAGATG 601  
 QY 733 AATTGCAATCGAACCCCAAGTGAAAAAATTTGCCGTAATTTGATCTGATGAGCT 792  
 DB 600 AATTGCAATCGAACCCCAAGTGAAAAAATTTGCCGTAATTTGATCTGATGAGCT 541  
 QY 793 GCAGTACACAGATGTTCTACCGCTCCACAAAGTGAAGATGTAATGTAATGTAATG 852  
 DB 540 GCAGTACACAGATGTTCTACCGCTCCACAAAGTGAAGATGTAATGTAATGTAATG 481  
 QY 853 ACTTTTATTTATTTCCCTGATCAAGACAGCTAATCTATTTTCAAGATGTTTAAAC 912  
 DB 480 ACTTTTATTTATTTCCCTGATCAAGACAGCTAATCTATTTTCAAGATGTTTAAAC 421  
 QY 913 CTTTGTGTGTGTGTATTAATAATGTTGTTATCTGTTGCTTCTGATACAGAC 972  
 DB 420 CTTTGTGTGTGTGTATTAATAATGTTGTTATCTGTTGCTTCTGATACAGAC 361  
 QY 973 TGTTCGCGGTGGTGTGAATATTTTGTGATGTTGATGTTATGATGATGATGATG 1032  
 DB 360 TGTTCGCGGTGGTGTGAATATTTTGTGATGTTGATGTTATGATGATGATGATG 301  
 QY 1033 GTGAGTTTATCTGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1092  
 DB 300 GTGAGTTTATCTGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 241  
 QY 1093 GTCTGTCTCTTCCATGATGATTAAGTTTGTGAGATGTTATGATGATGATGATG 1152  
 DB 240 GTCTGTCTCTTCCATGATGATTAAGTTTGTGAGATGTTATGATGATGATGATG 181  
 QY 1153 GTTCTCTCTTGTGAGATGTTAAATGGAAGATTAACATCTGATTAATGATGTTCT 1212  
 DB 180 GTTCTCTCTTGTGAGATGTTAAATGGAAGATTAACATCTGATTAATGATGTTCT 121  
 QY 1213 ACTGTGATATTAATTTGTGCTGAGAAATTTGTAATTTGTGCACTATGTAACAAA 1272  
 DB 120 ACTGTGATATTAATTTGTGCTGAGAAATTTGTAATTTGTGCACTATGTAACAAA 61  
 QY 1273 CAACGAGATATGTTAATAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAAT 1332  
 DB 60 CAACGAGATATGTTAATAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAAT 1

**RESULT 3**  
**LOCUS** AA746643 811 bp mRNA linear EST 22-JAN-1998  
**DEFINITION** nx27c08.s1 NCI CGAP GC4 Homo sapiens cDNA clone IMAGE:1257326 3' similar to gb:M28209 RAS-RELATED PROTEIN RAB-1A (HUMAN)., mRNA sequence.  
**ACCESSION** AA746643  
**VERSION** AA746643.1 GI:2786629  
**KEYWORDS** EST.  
**SOURCE** human.  
**ORGANISM** Homo sapiens  
**REFERENCE** 1 (base 1 to 811)  
**AUTHORS** NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
**JOURNAL** Unpublished (1997)  
**COMMENT** Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: www.bio.liml.gov/db/rdp/image/image.html  
 Insert length: 861 Std Error: 0.00  
 Seq primer: -40m13 fwd. Et from Amersham  
 High quality sequence stop: 471.  
 Location/Qualifiers  
 1. 811  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="1257326"  
 /clone\_lib="NCI-CGAP-GC4"  
 /tissue\_type="pooled germ cell tumors"  
 /lab\_host="DH10B"  
 /note="Vector: pRT30-Pac (Pharmacia) with a modified polylinker: 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - o19o(drf) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

**FEATURES**  
 source

BASE COUNT 269 a 165 c 145 g 232 t  
 ORIGIN

Query Match 51.8%; Score 727.6; DB 9; Length 811;  
 Best Local Similarity 95.8%; Pred. No. 9.8e-77;  
 Matches 780; Conservative 0; Mismatches 29; Indels 5; Gaps 3;

QY 502 AGGAATTCCTGATTCCTTGGAAATTCGTTTGGAAACCAAGCTGTAAGACACGA 561  
 DB 809 AGGAATTCCTGATTCCTTGGAAATTCGTTTGGAAACCAAGCTGTAAGACACGA 752  
 QY 562 ATGAGAACAGTCTTTCATGACGATGACGATGTAATTAAGCAATGGGTCGGAG 621  
 DB 751 ATGAGAACAGTCTTTCATGACGATGACGATGTAATTAAGCAATGGGTCGGAG 695  
 QY 622 CAACAGCTGTGTGCTGAGAGTCCATGTTAAATTTGAGACACTCAGTCAAGCAGT 681  
 DB 694 CAACAGCTGTGTGCTGAGAGTCCATGTTAAATTTGAGACACTCAGTCAAGCAGT 635  
 QY 682 CAGGTGAGGTGTGCTAAATTTTCCCTCCATCTTTTCTCAGCAATGATTTGCAA 741  
 DB 634 CAGGTGAGGTGTGCTAAATTTTCCCTCCATCTTTTCTCAGCAATGATTTGCAA 575  
 QY 742 TCTGAACCCAAAGTGAATAATTTGCTGAATTTGTAATGATGATGATGATGATG 801  
 DB 574 TCTGAACCCAAAGTGAATAATTTGCTGAATTTGTAATGATGATGATGATGATG 515  
 QY 802 CAGATTCCTAGCGTCTGACAGAGTGAATGTAATGTAATGTAATGTAATGTAATG 861  
 DB 514 CAGATTCCTAGCGTCTGACAGAGTGAATGTAATGTAATGTAATGTAATGTAATG 455  
 QY 862 TTATGCCCTGACCTCAAGACAGTCACTTCAAGTGAAGTGAAGTGAAGTGAAGT 921  
 DB 454 TTATGCCCTGACCTCAAGACAGTCACTTCAAGTGAAGTGAAGTGAAGTGAAGT 395  
 QY 922 CTGTTTATTAATTAATGTTGTAATCTGTTGCTTCTGTAACGAGCTGTTCCG 981  
 DB 394 CTGTTTATTAATTAATGTTGTAATCTGTTGCTTCTGTAACGAGCTGTTCCG 335  
 QY 982 TGGTGGTGAATTAATTTGTTGATGTTTAATGGAAGTGAAGTGAAGTGAAGT 1041  
 DB 334 TGGTGGTGAATTAATTTGTTGATGTTTAATGGAAGTGAAGTGAAGTGAAGT 275  
 QY 1042 AGTCTTCTGACATGAAGTTCAGCAATTTGTAATCAACAGCAAGCAAGCTGTCTGC 1101  
 DB 274 AGTCTTCTGACATGAAGTTCAGCAATTTGTAATCAACAGCAAGCAAGCTGTCTGC 215  
 QY 1102 TTTCATGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1161  
 DB 214 TTTCATGATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 155

QY	1162	TGTACAGCTTAAATGGAAGAGTACACTACATCTGATTTATAGTTCTTCATACCTCGAT	1221
DB	154	TGTACAGCTTAAATGGAAGAGTACACTACATCTGATTTATAGTTCTTCATACCTCGAT	95
QY	1222	ATAATTTGTGGCTGCAGATATTTGTAATTTGTTGCACACTATGTAAACAAACACTGAAG	1281
DB	94	ATAATTTGTGGCTGCAGATATTTGTAATTTGTTGCACACTATGTAAACAAACACTGAAG	35
QY	1282	ATATGTTTAAATTAATTTGTACTTATTTGGAGCTA	1315
DB	34	ATATGTTTAAATTAATTTGTACTTATTTGGAGCTA	1
RESULT 4			
LOCUS	B1869965	817 bp	mRNA
DEFINITION	603394116p1 NIH_MGC_90 Homo sapiens CDNA clone IMAGE:5404275 5',		
ACCESSION	B1869965		
VERSION	B1869965.1	GI:16043638	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgaqbs@remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LLM12030 row: p column: 04 High quality sequence stop: 669.		
FEATURES			
source	Location/Qualifiers		
	1..817		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5404275"		
	/clone_lib="NIH_MGC_90"		
	/tissue_type="adenocarcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: liver; Vector: PCMV-SpOrt6; site_1: Noti; site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.7 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library."		
BASE COUNT	237 a 155 c 169 g 256 t		
ORIGIN			
Query Match	51.5%; Score 723.8; DB 13; Length 817;		
Best Local Similarity	95.3%; Pred. No. 2.7e-76;		
Matches 779; Conservative	0; Mismatches 32; Indels 6; Gaps 3;		
QY	475	AGAAGTGTAGTACATACACACAGGAGGAAATTTGTCATCCCTTGGAATTCGCTTT	534
DB	1	AGAAGTGTAGTACATACACACAGGAGGAAATTTGTCATCCCTTGGAATTCGCTTT	60
QY	535	TGGAACACAGTGTCTAAGATGCAACGAATGTAGACAGCTTTTCATGACGATGCGACGTG	594
DB	61	TGGAACACAGTGTCTAAGATGCAACGAATGTAGACAGCTTTTCATGACGATGCGACGTG	120
QY	595	AGATTAAAAAGCGATGGGTCCCGAGCAACAGCTGTGTGCTGAGAGTCCAAATGTTA	654
DB	121	AGATTAAAAAGCGATGGGTCCCGAGCAACAGCTGTGTGCTGAGAGTCCAAATGTTA	180

FEATURES	LOCATION/Qualifiers
Y 655	AAATTCACAGACACTCCAGTCAAGCAGTACAGGTGAGGTGGCTGCTAAATATTTGCCCAT 714
Db 181	AAATTCACAGACACTCCAGTCAAGCAGTACAGGTGAGGTGGCTGCTAAATATTTGCCCAT 240
Qy 715	CCCTTTCTCAGACGAATGAATTTGCAATCTGTAACCAAGTGAACCAAAATTTGCCCTGA 774
Db 241	CCCTTTCTCAGACGAATGAATTTGCAATCTGTAACCAAGTGAACCAAAATTTGCCCTGA 300
Qy 775	ATTGACTGATGTATGCTGACACTACACAGATCTTACCCTGTCACCAAGTCAAGAT 834
Db 301	ATTGACTGATGTATGCTGACACTACACAGATCTTACCCTGTCACCAAGTCAAGAT 360
Qy 835	TGTAATAGTCAATCTGACCTTTTATTTTATTTCCCTGACACCAAGCTTACTTCAT 894
Db 361	TGTAATAGTCAATCTGACCTTTTATTTTATTTCCCTGACACCAAGCTTACTTCAT 420
Qy 895	TTTCAGAACTGTTTTAAACCTTTTGTCGTGTTTATTAATATGTGTATCCCTGTT 954
Db 421	TTTCAGAACTGTTTTAAACCTTTTGTCGTGTTTATTAATATGTGTATCCCTGTT 480
Qy 955	GCTTTCCTGATACACGACTGTTCCCGTGGTTGGTTAGAAATATATTTGTTGATGTT 1014
Db 481	GCTTTCCTGATACACGACTGTTCCCGTGGTTGGTTAGAAATATATTTGTTGATGTT 540
Qy 1015	AATATGGATGTTTATGATGTCAGGTTTATGCTCTTCGAAGATCAAGTTCAGCATTTGTA 1074
Db 541	AATATGGATGTTTATGATGTCAGGTTTATGCTCTTCGAAGATCAAGTTCAGCATTTGTA 600
Qy 1075	TCAACACACACAGACAGTGTCTGTACATTCATCATCAAAAGTTAGTACATTTAAT 1134
Db 601	TCAACACACACAGACAGTGTCTGTACATTCATCATCAAAAGTTAGTACATTTAAT 660
Qy 1135	GTAAGATCT---GATTTGCTAGTCTCTCTCTGTATGATTAATATGAAAGATTACACTA 1191
Db 661	GGAAGATCTGATTTGGCTAGTCTCTCTCTGTATGATTAATATGAAAGATTACACTA 720
Qy 1192	TCTGATTAATAGT-TTCTTCTTACCTGCTCAATATATTTGGCTGACAA--TATGTAA 1248
Db 721	TCTGATTAATAGTCTCTTCACTTCACTCTCATATATATGAGGGCTGCAGAACTTTCGCA 780
Qy 1249	TTTGTTCACACTATGTAAACAAACAACTGAAGATAT 1285
Db 781	TATGTGCCCACTATGTAAACAAACAACTGAAGATAT 817
RESULT 5	
BQ276678	910 bp mRNA linear EST 07-MAY-2002
LOCUS	BQ276678
DEFINITION	AGENCOURT_7025320 NIH_MGC_126 Homo sapiens cdna clone IMAGE:5809629
ACCESSION	BQ276678
VERSION	BQ276678.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 910)
TITLE	NIH-MGC http://mgs.nci.nih.gov/.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: NCI CDNA Library Preparation: Michael Brownstein Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLB) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB at: http://image.lnl.gov Plate: LNCM2055 row: 1 column: 22 High quality sequence stop: 558.

```

source
1. .910
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5809629"
/clone_lib="NIH_MGC_126"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/note="Vector: pDNR-LIB; Site.1: SfiI (ggccatctagcc);
Site.2: SfiI (ggccgcctcgcc); Double-stranded cDNA was
prepared from a pool of 40 cell line polyA+ RNAs (bladder
- 2%, blood - 3.4%, brain - 5.6%, breast - 12.5%, colon
- 4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
used in cloning as follows:
5'-AAGCAGTGTATCAGCAGAGTGGCCTTACGCGGG-3' and
5'-ATTCTAGAGCGCGGCGGCGGCGCATG-df(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5-1
kb size fraction (other fractions present in NIH_MGC_127
and NIH_MGC_128). Library created in the laboratory of T.
Uedlin, M.D., Ph.D. (NIH, NIH). Note: this is a NIH_MGC
Library"
BASE COUNT      243 a      194 c      180 g      292 t      1 others
ORIGIN
Query Match      50.4%; Score 708.8; DB 14; Length 910;
Best Local Similarity 98.9%; Pred. No. 1.4e-74;
Matches 724; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Oy 584 GATGCGACGTGAGTAAAGCGAATGGTCCCGGAGCAACAGCTGTGCTGAGAA 643
Db 3 GATGCGACGTGAGTAAAGCGAATGGTCCCGGAGCAACAGCTGTGCTGAGAA 62
Oy 644 GTCCAGTAAATTCAGAGCACTCCAGTCAAGAGTGAAGTGGTCTGTAAAA 703
Db 63 GTCCAGTAAATTCAGAGCACTCCAGTCAAGAGTGAAGTGGTCTGTAAAA 122
Oy 704 TTTGGCTCATCCCTTTTCACAGCAATGAAATTTGCAATCGAACCCAGTAAAAA 763
Db 123 TTTGGCTCATCCCTTTTCACAGCAATGAAATTTGCAATCGAACCCAGTAAAAA 182
Oy 764 AAATTCCTGAATTTGACTGTATGACTGCACATACACAAATTCCTTACCGTCCACAA 823
Db 183 AAATTCCTGAATTTGACTGTATGACTGCACATACACAAATTCCTTACCGTCCACAA 242
Oy 824 AGTCAGAGATGTGAATGCTCAATGCTGATGCTGCTGCTGCTGCTGCTGCTG 883
Db 243 AGTCAGAGATGTGAATGCTCAATGCTGATGCTGCTGCTGCTGCTGCTGCTG 302
Oy 884 CTAATCTTATTTTCAAGACTGTTTAAACCTTTGTGTGCTGTTTAAATTAATGCTG 943
Db 303 CTAATCTTATTTTCAAGACTGTTTAAACCTTTGTGTGCTGTTTAAATTAATGCTG 362
Oy 944 TAAATCTTGTGCTTCTGATACAGAGCTGTTCCCGGTGGTTAGAAATTAATTTG 1003
Db 363 TAAATCTTGTGCTTCTGATACAGAGCTGTTCCCGGTGGTTAGAAATTAATTTG 422
Oy 1004 TTTTATGATTTAATTTGAGATGCTGATGCTGATGCTGATGCTGATGCTGATG 1063
Db 423 TTTTATGATTTAATTTGAGATGCTGATGCTGATGCTGATGCTGATGCTGATG 482
Oy 1064 GCCATTTTGTATCAAAACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1123
Db 483 GCCATTTTGTATCAAAACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542
Oy 1124 AGATGTTATTTATGATGCTGATTTGCTAGTCTTCTGTTAGAGTTATTAATGAAGA 1183
Db 543 AGATGTTATTTATGATGCTGATTTGCTAGTCTTCTGTTAGAGTTATTAATGAAGA 602
Oy 1184 TTACGCTATCTGATTAATAGTTTCTCATCTGCTGATTAATTTGTGGCTGAGAAAT 1243

```

```

Db 603 TTACACTATCTGATTAATAGTTTCTTCACTCCGATATATAATTTGTGCTGCACAAAT 662
Oy 1244 TGTAAATTTGTTCACACATATGATTAACAAACAGTGAAGATA-TGTTAATAAATTTGTA 1302
Db 663 TGTAAATTTGTTCACACATATGATTAACAAACAGTGAAGATA-TGTTAATAAATTTGTA 722
Oy 1303 CTTATGGAAGT 1314
Db 723 CTTATGGAAT 734

RESULT 6
BO772048/c
LOCUS
DEFINITION
UI-H-E21-bdk-1-10-0-UI-s1 NCI CGAP CH2 Homo sapiens cDNA clone
BO772048
UI-H-E21-bdk-1-10-0-UI 3', mRNA sequence.
BO772048.1 GI:21980524
VERSION
KEYWORDS
EST
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 761)
NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Dr. Steven Gitelis/ Rush Presbyterian, Dept. of
Orthopedics
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@iowa.edu
Seq primer: M13 FORWARD
POLYA-Yes.

FEATURES
location/Qualifiers
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-H-E21-bdk-1-10-0-UI"
/clone_lib="NCI CGAP CH2"
/tissue_type="Chondrosarcoma Grade II"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Organ: Left Pelvis; Vector: pT73-Pac (Pharmacia)
with a modified polylinker; Site.1: EcoR I; Site.2: Not I;
NCI CGAP CH2 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma Grade II. The library
was constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
TGATCGGCT.
TAG-LIB-UI-H-E21
TAG-TISSUE-Grade-2-chondrosarcoma
TAG-SEQ-ATCTAATATG"
BASE COUNT      258 a      143 c      125 g      235 t
ORIGIN
Query Match      50.4%; Score 708.4; DB 14; Length 761;
Best Local Similarity 99.3%; Pred. No. 1.8e-74;
Matches 753; Conservative 0; Mismatches 1; Indels 4; Gaps 4;
Oy 575 TTTGATGAGATGAGTGAATTAAGGATGGTCCGGAGCAACAGCTGTG 634

```



```

Oy 684 -GTGAGAGTGTCTGCTAAATTTGGCCATCCCTTTTCTCACAGCAATGAATTTGCAAT 742
      |||
Db 720 GGTGTGAGGTGTGCTGTAATTTTGGCTCCATCCTTTTCTCACAGCAATGAATTTGCAAT 779
Oy 743 CTGAACCCAGTAGTAAACAAATTTGCCGTAATGTACTATAGTGCATACAC 802
      |||
Db 780 CTGAACCCAGTAGTAAACAAATTTGCCGTAATGTACTATAGTGCATACAC 839
Oy 803 AGATTCTTACCGTCTCCACAAAGCTCAGAGATTGTAATGCTCAATAGCTTTT 862
      |||
Db 840 AGATTCTTACCGTCTCCACAAAGCTCAGAGATTGTAATGCTCAATAGCTTTT 899
Oy 863 TATTCCTTACTCAGACAGCTACTTCTTTTTCAGAACTGTT 907
      |||
Db 900 TATTCCTTACTCAGACAGCTACTTCTTTTTCAGAACTGTT 944

RESULT 8
BM452262 1105 bp mRNA linear EST 05-FEB-2002
LOCUS BM452262
DEFINITION AGENCOURT_6386191 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5526635
ACCESSION BM452262
VERSION BM452262.1 GI:18501302
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgaapb-remail.nih.gov
Tissue Procurement: ATCC/DCMD/DFP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM12200 row: j column: 12
High quality sequence stop: 733.
Location/Qualifiers
1. 1105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5526635"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/label="Organ: skin; Vector: pCMV-Sport6; Site: 1: NotI;
Site: 2: SalI; Cloned unidirectionally; primer: oligo df.
Average insert size 2 kb. Library constructed by Life
Technologies."
BASE COUNT 320 a 216 c 279 g 282 t 8 others
ORIGIN
Query Match 50.1%; Score 704.4; DB 13; Length 1105;
Best Local Similarity 86.0%; Pred. No. 4.1e-74;
Matches 899; Conservative 0; Mismatches 43; Indels 103; Gaps 7;

Oy 1 AAGCGATAGCTGACT- GCGGCGGCTGCTGATTTGTTCTAGAGGAGCGATAGGGAGA 59
      |||
Db 2 AAGCGATAGCTGAGTGGCGGCTGCTGATTTGTTCTAGAGGAGCGATAGGGAGA 61
Oy 60 CGTTTGTCTCCCGGAACGCTTATCTCATCTTTCTTTTGATTAACCCGCGCGGAG 119
      |||
Db 62 CGTTTGTCTCCCGGAACGCTTATCTCATCTTTCTTTTGATTAACCCGCGCGGAG 121
Oy 120 AGTCAGGCGCGGCTGCGGACGACGAAGCGGCGGTGCGCGCGGCGGAGTCAGTGAC 179
      |||

Db 122 AGTCAGGCGCGGCTGCGGACGACGAAGCGGCGGTGCGCGCGGCGGAGTCAGTGAC 181
Oy 180 ATGTCCAGATGAATCCCGAATATATTTATTTATTCAGTACTTGTATGGCGACTCA 239
      |||
Db 182 ATGTCCAGATGAATCCCGAATATATTTATTTATTCAGTACTTGTATGGCGACTCA 241
Oy 240 GGGGTGGAAGTCTTCCCTTCTCTTGTAGTTTCAGATGATATATACAGAAACTAC 299
      |||
Db 242 GGGGTGGAAGTCTTCCCTTCTCTTGTAGTTTCAGATGATATATACAGAAACTAC 301
Oy 300 ATCAGACAAATTTGGTGTGATTTCAAAATTAAGACTATAGTGAACGGGAACATAC 359
      |||
Db 302 ATCAGACAAATTTGGTGTGATTTCAAAATTAAGACTATAGTGAACGGGAACATAC 361
Oy 360 AAGCTTCAAAATTTGGTGTGATTTCAAAATTAAGACTATAGTGAACGGGAACATAC 371
      |||
Db 362 AAGCTTCAAAATTTGGTGTGATTTCAAAATTAAGACTATAGTGAACGGGAACATAC 421
Oy 372 -----GAGTCTTCAT 383
      |||
Db 422 TACAGAGACGCCCATGGCATCATAGTTGTGTATGATGTACAGATCAGAGAGTCTTCAT 481
Oy 384 AATGTTAAACAGTGGCTGACGAGAAATAGATGTTATGCCAGTGAATGTCAACAAATG 443
      |||
Db 482 AATGTTAAACAGTGGCTGACGAGAAATAGATGTTATGCCAGTGAATGTCAACAAATG 541
Oy 444 TTGTTAGGGAACAAATGTGATCTGACCAACAAAGAAAGTAGTACTACACAGCGAAG 503
      |||
Db 542 TTGTTAGGGAACAAATGTGATCTGACCAACAAAGAAAGTAGTACTACACAGCGAAG 601
Oy 504 GAATTTGCTGATTCCTTTGGAAATTCGCTTTTGGAAACGAGTGAAGAAATGCAACGAT 563
      |||
Db 602 GAATTTGCTGATTCCTTTGGAAATTCGCTTTTGGAAACGAGTGAAGAAATGCAACGAT 661
Oy 564 GTGAACAGTCTTTTCATGACAGATGCGACCTGATTAACAAACGAAATGGTCCCGGAGCA 623
      |||
Db 662 GTGAACAGTCTTTTCATGACAGATGCGACCTGATTAACAAACGAAATGGTCCCGGAGCA 721
Oy 624 ACAGCTGTGTGCTGGAAGTCAATGTTAAATTCAGAGCAGTCAAGCAGCTCA 683
      |||
Db 722 ACAGCTGTGTGCTGGAAGTCAATGTTAAATTCAGAGCAGTCAAGCAGCTCA 780
Oy 684 -GGTGAGGTGCTGCTAAATTTTGCCTCCATCCTTTTCTCACAGCAATGAATTTGCAAT 742
      |||
Db 781 NGGTGAGGTGCTGCTAAATTTTGCCTCCATCCTTTTCTCACAGCAATGAATTTGCAAT 840
Oy 743 CTGAACCCAGTAGTAAACAAATTTGCCGTAATGTACTATAGTGCATACAC 802
      |||
Db 841 TGAACCCAGTAGTAAACAAATTTGCCGTAATGTACTATAGTGCATACAC 899
Oy 803 AGATTCTTACCGTCTCCACAAAGCTCAGAGATTGTAATGCTCAATAGCTTT 860
      |||
Db 900 AGATTCTTACCGTCTCCACAAAGCTCANNAGATTAATGGCAATAGCTGACTTTT 959
Oy 861 TTTATTCCTTGAATCAAGACAGCTAATCTTATTTTCAACACTGTTTAAAC- TTTG 919
      |||
Db 960 TTTATTCCTTGAATCAAGACAGCTAATCTTATTTTCAACACTGTTTAAAC- TTTG 1019
Oy 920 TGTGTGTTTAAATATATGTGT 944
      |||
Db 1020 TGTGTGTTTAAATATATATGTGT 1044

RESULT 9
BQ448090/c 725 bp mRNA linear EST 29-MAY-2002
LOCUS BQ448090
DEFINITION UI-H-B01-bal-k-14-0-UI.s1 NCI-60AP-CE1 Homo sapiens cDNA clone
ACCESSION BQ448090
VERSION BQ448090.1 GI:21251202
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```





```

|||||
Db 565 TGCGTGAATGATCTGATGAGCTGCACACACAGATCTTACCGCTCCACAAAGGT 506
QY 828 CAGAGATGTAATGGCACTACTGACTTTTATTTATCCCTGACACAGACAGCTAA 887
Db 505 CAGAGATGTAATGGCACTACTGACTTTTATTTATCCCTGACACAGACAGCTAA 446
QY 888 CTTCATTTTTCAGACAGCTGTTTAAACCTTGTGTCGCTGTTTAAATTAATGCTGTAA 947
Db 445 CTTCATTTTTCAGACAGCTGTTTAAACCTTGTGTCGCTGTTTAAATTAATGCTGTAA 386
QY 948 CCTGTGCTTCTCTGATACAGACGCTTTCCCTGCTGCTGTTAGATATATTTGTTTT 1007
Db 385 CCTGTGCTTCTCTGATACAGACGCTTTCCCTGCTGCTGTTAGATATATTTGTTTT 326
QY 1008 GATGTTATATGTCATGTTAGATGTCAGGTTAGTCTGTCGAGATGAAGTGCAGCA 1067
Db 325 GATGTTATATGTCATGTTAGATGTCAGGTTAGTCTGTCGAGATGAAGTGCAGCA 266
QY 1068 TTTTGTATCAAAACGACACAGACAGCTGCTGTCACCTTCCATGCAATAAAGTTAGTAGAT 1127
Db 265 TTTTGTATCAAAACGACACAGACAGCTGCTGTCACCTTCCATGCAATAAAGTTAGTAGAT 206
QY 1128 GTTATATGTAAGATCTGATTTGCTAGTCTTCTCTGTTAGAGTTTAAATGGAAGATTAC 1187
Db 205 GTTATATGTAAGATCTGATTTGCTAGTCTTCTCTGTTAGAGTTTAAATGGAAGATTAC 146
QY 1188 ACATATGATTAATAGTTTCTTCATACCTGTCATATATATTTGGCTGCGAGATATGTA 1247
Db 145 ACATATGATTAATAGTTTCTTCATACCTGTCATATATATTTGGCTGCGAGATATGTA 86
QY 1248 ATTTGTTGCACACTATGTAAACAAACAGTGAAGATATGTTTAAATATATTTACTTAT 1307
Db 85 ATTTGTTGCACACTATGTAAACAAACAGTGAAGATATGTTTAAATATATTTACTTAT 26
QY 1308 TGGAACTAAAAAAGAAAAA 1332
Db 25 TGGAACTAAAAAAGAAAAA 1

RESULT 12
BM919860
LOCUS 976 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6708189 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5749635
ACCESSION BM919860
VERSION 5, mRNA sequence.
KEYWORDS BM919860.1 GI:19370239
SOURCE EST.
ORGANISM Homo sapiens
            human.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgabs-remail.nih.gov
          Tissue Procurement: Life Technologies, Inc.
          CDNA Library Preparation: Life Technologies, Inc.
          CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LNL at:
          http://image.llnl.gov
          Plate: LHAM12779 row: f column: 04
          High quality sequence stop: 757.
          Location/Qualifiers
            1..976
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5749635"
            /clone_lib="NIH_MGC_120"

```

```

/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
pcmv-SPOrt6; Site.1: NotI; Site.2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT 285 a 196 c 246 g 248 t 1 others
ORIGIN
Query Match 49.2%; Score 691.8; DB 14; Length 976;
Best Local Similarity 87.6%; Pred. No. 1.3e-72;
Matches 832; Conservative 0; Mismatches 18; Indels 100; Gaps 3;

QY 16 GCGCGGCGCTGCTGATTTGTGTTCTAGGGAGCGAGTACGGGAACGTTGCTCCCGGA 75
Db 13 GCGCGGCGCTGCTGATTTGTGTTCTAGGGAGCGAGTACGGGAACGTTGCTCCCGGA 72
QY 76 ACAGCTATGTCATCTTCTTCTGATTTACCGCTGGCGGAGAGTCAGGGCGGCGCT 135
Db 73 ACAGCTATGTCATCTTCTTCTGATTTACCGCTGGCGGAGAGTCAGGGCGGCGCT 132
QY 136 GCGGCGACGAAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 195
Db 133 GCGGCGACGAAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 192
QY 196 CCGAATATGATTTATTTTCAAGTTACTTCTGATTTGCGGCACTCAGGGGTGGAAGTCTT 255
Db 193 CCGAATATGATTTATTTTCAAGTTACTTCTGATTTGCGGCACTCAGGGGTGGAAGTCTT 252
QY 256 GCGTCTTCTTAGGTTTGCAGATGATATATACAGAACTCATAGCAATGAGCAATGGTG 315
Db 253 GCGTCTTCTTAGGTTTGCAGATGATATATACAGAACTCATAGCAATGAGCAATGGTG 312
QY 316 TGGATTTCAAAATTAAGACTATAGAGTTAGACGGGAACAATCAACCTTCAATA---- 371
Db 313 TGGATTTCAAAATTAAGACTATAGAGTTAGACGGGAACAATCAACCTTCAATAATGGG 372
QY 372 ----- 371
Db 373 ACACAGCAGGCCAGGAAGATTTGCAACATCACTCTAGTTATTACAGAGGCCCATG 432
QY 372 -----GAGTCCCTCAATTAATGTTAAACAGTGGC 399
Db 433 GCATCATAGTTGTTGATGATGTGACAGATCAGAGAGTCTTCAATTAATGTTAAACAGTGGC 492
QY 400 TGCAGGAATAGATCGTTATGCGCAGTGAATAATGTTCAAAATTTGTTAGGGAACAAAT 459
Db 493 TGCAGGAATAGATCGTTATGCGCAGTGAATAATGTTCAAAATTTGTTAGGGAACAAAT 552
QY 460 GTGATCTGACCACAAAGAGTACTAGACTACACACAGCGAAGAAATTTGCTGATTTCC 519
Db 553 GTGATCTGACCACAAAGAGTACTAGACTACACACAGCGAAGAAATTTGCTGATTTCC 612
QY 520 TTGGAATTCGTTTGTGGAACCGAGTGTGAAGATGCAAGAAATGTAAGACAGTCTTCA 579
Db 613 TTGGAATTCGTTTGTGGAACCGAGTGTGAAGATGCAAGAAATGTAAGACAGTCTTCA 672
QY 580 TGAAGATGAGCAGTGAATTAAGGAATGGTTCGCGGAGCAACAGCTGTTGTTGCTG 639
Db 673 TGAAGATGAGCAGTGAATTAAGGAATGGTTCGCGGAGCAACAGCTGTTGTTGCTG 732
QY 640 AGAAGTCAATGTTAAATTCAGAGCACTCCAGTCAAGCACTCAGGTGAGGTGCTGCT 699
Db 733 AGAAGTCAATGTTAAATTCAGAGCACTCCAGTCAAGCACTCAGGTGAGGTGCTGCT 792
QY 700 AAAATTTGCTCCATCTTCTTCTCAGAGCAATGAAATTTGAATCTGAACCCCAAGTGA 759
Db 793 AAAATTTGCTCCATCTTCTTCTCAGAGCAATGAAATTTGAATCTGAACCCCAAGTGA 852

```



Yy	760	AAACAAATTCGCAATTCGTAAGTACGACGACAAAG-ATTCTTACCCGTC	818
Db	853	AAACAAATTCGCAATTCGTAAGTACGACGACAAAG-ATTCTTACCCGTC	912
Yy	819	CACAAAGTCAGAGA---TGTAAATGTCATPACTGCTTTTTTAT	865
Db	913	CACAAAGTCAGAGAATTTGTAATGTCATPACTGACCTTTTTTTT	962
RESULT 13			
LOCUS	B1086445	837 bp	mRNA
DEFINITION	602849703f1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:491233 5',		
ACCESSION	B1086445		
VERSION	B1086445.1	GI:14504775	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 837)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: c9apbs-r@mail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: Incyte Genomics, Inc.		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLML1008 row: f column: 02		
FEATURES	High quality sequence stop: 758.		
SOURCE	Location/Qualifiers		
	1..837		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone_image="491233"		
	/clone_lib="NIH_MGC_10"		
	/cell_line="MGC36"		
	/lab_host="DH10B"		
	/note="Organ: Cervix; Vector: pCMV-Sport6; Site_1: Nctli;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.		
	Average insert size 1.5 kb. Library prepared by Life		
	Technologies."		
BASE COUNT	248 a 134 c 174 g 281 t		
ORIGIN			
Query Match	49.2%; Score 690.6; DB 13; Length 837;		
Best Local Similarity	97.5%; Pred. No. 2.1e-72;		
Matches 754; Conservative	0; Mismatches 14; Indels 5; Gaps 5;		
Yy	557	AACGATGTGAACAGCTTTCATGACGATGCGACGTAGATTAACAAAGCATGGCTCC	616
Db	1	AACGATGTGAACAGCTTTCATGACGATGCGACGTAGATTAACAAAGCATGGCTCC	60
Yy	617	CGGACCAACAGCTGGTGGTGGTGAAGTCCATGTAAATTCAGACATCCAGTCAA	676
Db	61	CGGACCAACAGCTGGTGGTGGTGAAGTCCATGTAAATTCAGACATCCAGTCAA	120
Yy	677	GGAGTCAGGAGGTGCTGCTAAATTTGCTCCATCTTTCTCCACACATGAATT	736
Db	121	GGAGTCAGGAGGTGCTGCTAAATTTGCTCCATCTTTCTCCACACATGAATT	180
Yy	737	TGCATCTGAACCAAGTCAAAATTCCTGATATTGACTGATGTAGCTGCAC	796
Db	181	TGCATCTGAACCAAGTCAAAATTCCTGATATTGACTGATGTAGCTGCAC	240
Yy	797	TACACAGATTCCTACGCTCCACAAAGTCAGAGATTGTAATGTCATPACTG	856
Yy	241	TACACAGATTCCTACGCTCCACAAAGTCAGAGATTGTAATGTCATPACTG	300
Yy	857	TTTTTTATTCCTTGACCTCAAGACAGTAACTTATTCAGAACTTTTAACTTT	916
Db	301	TTTTTTATTCCTTGACCTCAAGACAGTAACTTATTCAGAACTTTTAACTTT	360
Yy	917	GTCGCGGTGTTTAAATTAATGTCGTAATTCCTGCTGCTGCTGATACAGACGCT	976
Db	361	GTCGCGGTGTTTAAATTAATGTCGTAATTCCTGCTGCTGCTGATACAGACGCT	420
Yy	977	TCCCGTGGTGGTTAATATATTTGTTTGGATGTTTATATGCGATGTTTAACTGA	1036
Db	421	TCCCGTGGTGGTTAATATATTTGTTTGGATGTTTATATGCGATGTTTAACTGA	480
Yy	1037	GCTTATAGTCTTCTGAAAGTGAAGTCAAGCCATTTTGTATCAACAGACAGCA-GTGTC	1095
Db	481	GCTTATAGTCTTCTGAAAGTGAAGTCAAGCCATTTTGTATCAACAGACAGCA-GTGTC	540
Yy	1096	TGTCACCTTCCATGATTAAGTGTGATGATGTTATGATGAATCTGATTTGCTAGT	1155
Db	541	TGTCACCTTCCATGATTAAGTGTGATGATGTTATGATGAATCTGATTTGCTAGT	600
Yy	1156	CTTCCTTGTAGAGTTAT-AAATGGAAGAAATTAAC-TATCGATTAATAGTTCTCAT	1213
Db	601	CTTCCTTGTAGAGTTAT-AAATGGAAGAAATTAAC-TATCGATTAATAGTTCTCAT	660
Yy	1214	CTTCGATTAATTTTGGTGGTGGTCAATATGTAATTTTGGACACTATGTAACAAAC	1273
Db	661	CTTCGATTAATTTTGGTGGTGGTCAATATGTAATTTTGGACACTATGTAACAAAC	719
Yy	1274	AACGAAAGATGTTTAAATTAATGTAATTTGTAATTTGGAAGTAAATTAATTAAT	1326
Db	720	AA-TGAAGTATGTTTAAATTAATGTAATTTGTAATTTGGAAGTAAATTAATTAAT	771
RESULT 14			
LOCUS	BM706159	689 bp	mRNA
DEFINITION	UI-E-DW0-agg-f-10-0-UI-r1 UI-E-DW0 Homo sapiens cDNA clone		
ACCESSION	UI-E-DW0-agg-f-10-0-UI 5', mRNA sequence.		
VERSION	BM706159.1	GI:19019417	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 689)		
	Bonaldi,M.F., Lennon,G. and Soares,M.B.		
	Normalization and subtracgion: two approaches to facilitate gene		
	discovery		
	Genome Res. 6 (9), 791-806 (1996)		
	97044477		
JOURNAL	CONTACT: Soares, MB		
MEDLINE	Program for Rat Gene Discovery and Mapping		
COMMENT	University of Iowa		
	451 Eckstein Medical Research Building Iowa City, IA 52242, USA		
	Tel: 319 335 8250		
	Fax: 319 335 9565		
	Email: msotares@iui.uiowa.edu		
	tissue procurement: Dr. Gregg Hageman		
	cDNA library preparation: Dr. M. Bento Soares, University of Iowa		
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa		
	DNA Sequencing by:		

```

/tissue_type="lens"
/dev_stage="adult"
/lat_host="BDH10B (Life Technologies) (T1 phage resistant)"
/notes="organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UT-E-DN0 is a cDNA library containing the following
tissue(s): lens. The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (GT)18 tail. The
sequence tag for this library is CGATTACGCA. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

```

Query Match	49.0%;	Score 689;	DB 14;	Length 689;
Best Local Similarity	100.0%;	Pred. No. 3,6e-72;		
Matches 689;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 503	GGAAATTCGTGATTCCTTGGAATTCGGATTTCTTTTGGAAACCACTGCTAAGAAATGCAACGAA	562		
Db 1	GGAAATTCGTGATTCCTTGGAATTCGGATTTCTTTTGGAAACCACTGCTAAGAAATGCAACGAA	60		
QY 563	TGTAGAACAGTCCTTCATAGCAGATGGCAGCTGAGATTAAAAAGCAATGGTCCGGAGC	622		
Db 61	TGTAGAACAGTCCTTCATAGCAGATGGCAGCTGAGATTAAAAAGCAATGGTCCGGAGC	120		
QY 623	AACAGCTGGTGGTGTGAGAAAGTCCAAATGTTAAATTCAGACACTCCGATCAACAGTC	682		
Db 121	AACAGCTGGTGGTGTGAGAAAGTCCAAATGTTAAATTCAGACACTCCGATCAACAGTC	180		
QY 683	AGGTGAGAGTGGCTCTAAATTTGCTTCATCCTTTCTCTCAGCAATGAATTTGCAT	742		
Db 181	AGGTGAGAGTGGCTCTCTAAATTTGCTTCATCCTTTCTCTCAGCAATGAATTTGCAT	240		
QY 743	CTGAACCCCAAGGAAAAAATAATGGCTGAATGTGTAATGTATGATGACATACAC	802		
Db 241	CTGAACCCCAAGGAAAAAATAATGGCTGAATGTGTAATGTATGATGACATACAC	300		
QY 803	AGATCTTACCGCTCCACAAAGGTCAAGATTTGAAATGTCATACGACTTTT	862		
Db 301	AGATCTTACCGCTCCACAAAGGTCAAGATTTGAAATGTCATACGACTTTT	360		
QY 863	TATTCCTTGACTCAGACAGCTAACTTCATTTTCAGAACTGTATTTAACTTTGTGC	922		
Db 361	TATTCCTTGACTCAGACAGCTAACTTCATTTTCAGAACTGTATTTAACTTTGTGC	420		
QY 923	TGGTTATTAATAATATGATGTAATCTGTGCTTCTCTGATTAACACAGCTTTCCGT	982		
Db 421	TGGTTATTAATAATATGATGTAATCTGTGCTTCTCTGATTAACACAGCTTTCCGT	480		
QY 983	GGTGGTAGAATATATTTTGTTTGATGTTTAAATTTGGCAGTTTAGATGTCAGATT	1042		
Db 481	GGTGGTAGAATATATTTTGTTTGATGTTTAAATTTGGCAGTTTAGATGTCAGATT	540		
QY 1043	GTCCTTCTGAAGATGAAGTTCACGCACTTTTGTATCAACAGCAAGCAGATGCTGTCACT	1102		
Db 541	GTCCTTCTGAAGATGAAGTTCACGCACTTTTGTATCAACAGCAGATGCTGTCACT	600		
QY 1103	TTCCATGCATAAAGTTTATGATGATGTATATGTAAAGATCTGATTTGCTAGTCTTCTT	1162		
Db 601	TTCCATGCATAAAGTTTATGATGATGTATATGTAAAGATCTGATTTGCTAGTCTTCTT	660		
QY 1163	GTAGAGTATTAATGGAAGATTACACTA 1191			
Db 661	GTAGAGTATTAATGGAAGATTACACTA 689			

REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
BI092003				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
JOURNAL				
COMMENT				

BI092003 737 bp mRNA Linear EST 20-JUN-2001  
 602856302F1 NIH\_MGC\_10 Homo sapiens cDNA clone IMAGE:4997688 5',  
 mRNA sequence.  
 BI092003  
 BI092003.1 GI:14510333  
 EST.  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 737)  
 NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Sequencing Arrayed by: Incyte Genomics, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNT at:  
<http://image.llnl.gov>  
 plate: LLAM1025 row: C column: 01  
 High quality sequence, stop: 727.

FEATURES	Location/Qualifiers
source	1. .737

Plate: LLAM11025 row: C column: 01  
High quality sequence stop: 727.

Location/Qualifiers  
1. .737

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"
```

```
/clone="IMAGE:4997688"  
/clone_lib="NIH_MGC_10"
```

```
/cell_line="MGC36"  
/lab_host="DH10B"
```

```

/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

```

Average insert size 1.5 kb. Library prepared by Life Technologies."

214 a 135 c 147 g 241 t

49.0%: score 688.2: DB 13: length 737:

	1	2	3
Similarity	97.8%	97.8%	97.8%
Conservative	97.8%	97.8%	97.8%
Mismatches	13	13	13
Indels	3	3	3
Gaps	2	2	2

Conservative	Mismatches	Indels	gaps
13/	13/	3/	2/

50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526  
 527  
 528  
 529  
 530  
 531  
 532  
 533  
 534  
 535  
 536  
 537  
 538  
 539  
 540  
 541  
 542  
 543  
 544  
 545  
 546  
 547  
 548  
 549  
 550  
 551  
 552  
 553  
 554  
 555  
 556  
 557  
 558  
 559  
 560  
 561  
 562  
 563  
 564  
 565  
 566  
 567

550  
 551  
 552  
 553  
 554  
 555  
 556  
 557  
 558  
 559  
 560  
 561  
 562  
 563  
 564  
 565  
 566  
 567  
 568  
 569  
 570  
 571  
 572  
 573  
 574  
 575  
 576  
 577  
 578  
 579  
 580  
 581  
 582  
 583  
 584  
 585  
 586  
 587  
 588  
 589  
 590  
 591  
 592  
 593  
 594  
 595  
 596  
 597  
 598  
 599  
 600  
 601  
 602  
 603  
 604  
 605  
 606  
 607  
 608  
 609  
 610  
 611  
 612  
 613  
 614  
 615  
 616  
 617  
 618  
 619  
 620  
 621  
 622  
 623  
 624  
 625  
 626  
 627  
 628  
 629  
 630  
 631  
 632  
 633  
 634  
 635  
 636  
 637  
 638  
 639  
 640  
 641  
 642  
 643  
 644  
 645  
 646  
 647  
 648  
 649  
 650  
 651  
 652  
 653  
 654  
 655  
 656  
 657  
 658  
 659  
 660  
 661  
 662  
 663  
 664  
 665  
 666  
 667  
 668  
 669  
 670  
 671  
 672  
 673  
 674  
 675  
 676  
 677  
 678  
 679  
 680  
 681  
 682  
 683  
 684  
 685  
 686  
 687  
 688  
 689  
 690  
 691  
 692  
 693  
 694  
 695  
 696  
 697  
 698  
 699  
 700  
 701  
 702  
 703  
 704  
 705  
 706  
 707  
 708  
 709  
 710  
 711  
 712  
 713  
 714  
 715  
 716  
 717  
 718  
 719  
 720  
 721  
 722  
 723  
 724  
 725  
 726  
 727  
 728  
 729  
 730  
 731  
 732  
 733  
 734  
 735  
 736  
 737  
 738  
 739  
 740  
 741  
 742  
 743  
 744  
 745  
 746  
 747  
 748  
 749  
 750  
 751  
 752  
 753  
 754  
 755  
 756  
 757  
 758  
 759  
 760  
 761  
 762  
 763  
 764  
 765  
 766  
 767  
 768  
 769  
 770  
 771  
 772  
 773  
 774  
 775  
 776  
 777  
 778  
 779  
 780  
 781  
 782  
 783  
 784  
 785  
 786  
 787  
 788  
 789  
 790  
 791  
 792  
 793  
 794  
 795  
 796  
 797  
 798  
 799  
 800  
 801  
 802  
 803  
 804  
 805  
 806  
 807  
 808  
 809  
 810  
 811  
 812  
 813  
 814  
 815  
 816  
 817  
 818  
 819  
 820  
 821  
 822  
 823  
 824  
 825  
 826  
 827  
 828  
 829  
 830  
 831  
 832  
 833  
 834  
 835  
 836  
 837  
 838  
 839  
 840  
 841  
 842  
 843  
 844  
 845  
 846  
 847  
 848  
 849  
 850  
 851  
 852  
 853  
 854  
 855  
 856  
 857  
 858  
 859  
 860  
 861  
 862  
 863  
 864  
 865  
 866  
 867  
 868  
 869  
 870  
 871  
 872  
 873  
 874  
 875  
 876  
 877  
 878  
 879  
 880  
 881  
 882  
 883  
 884  
 885  
 886  
 887  
 888  
 889  
 890  
 891  
 892  
 893  
 894  
 895  
 896  
 897  
 898  
 899  
 900  
 901  
 902  
 903  
 904  
 905  
 906  
 907  
 908  
 909  
 910  
 911  
 912  
 913  
 914  
 915  
 916  
 917  
 918  
 919  
 920  
 921  
 922  
 923  
 924  
 925  
 926  
 927  
 928  
 929  
 930  
 931  
 932  
 933  
 934  
 935  
 936  
 937  
 938  
 939  
 940  
 941  
 942  
 943  
 944  
 945  
 946  
 947  
 948  
 949  
 950  
 951  
 952  
 953  
 954  
 955  
 956  
 957  
 958  
 959  
 960  
 961  
 962  
 963  
 964  
 965  
 966  
 967  
 968  
 969  
 970  
 971  
 972  
 973  
 974  
 975  
 976  
 977  
 978  
 979  
 980  
 981  
 982  
 983  
 984  
 985  
 986  
 987  
 988  
 989  
 990  
 991  
 992  
 993  
 994  
 995  
 996  
 997  
 998  
 999  
 1000

33653536373839404142434445464748495051525354555657585960616263646566676869707172737475767778798081828384858687888990919293949596979899100101102103104105106107108109110111112113114115116117118119120121122123124125126127128129130131132133134135136137138139140141142143144145146147148149150151152153154155156157158159160161162163164165166167168169170171172173174175176177178179180181182183184185186187188189190191192193194195196197198199200201202203204205206207208209210211212213214215216217218219220221222223224225226227228229230231232233234235236237238239240241242243244245246247248249250251252253254255256257258259260261262263264265266267268269270271272273274275276277278279280281282283284285286287288289290291292293294295296297298299300301302303304305306307308309310311312313314315316317318319320321322323324325326327328329330331332333334335336337338339340341342343344345346347348349350351352353354355356357358359360361362363364365366367368369370371372373374375376377378379380381382383384385386387388389390391392393394395396397398399400401402403404405406407408409410411412413414415416417418419420421422423424425426427428429430431432433434435436437438439440441442443444445446447448449450451452453454455456457458459460461462463464465466467468469470471472473474475476477478479480481482483484485486487488489490491492493494495496497498499500501502503504505506507508509510511512513514515516517518519520521522523524525526527528529530531532533534535536537538539540541542543544545546547548549550551552553554555556557558559560561562563564565566567568569570571572573574575576577578579580581582583584585586587588589590591592593594595596597598599600601602603604605606607608609610611612613614615616617618619620621622623624625626627628629630631632633634635636637638639640641642643644645646647648649650651652653654655656657658659660661662663664665666667668669670671672673674675676677678679680681682683684685686687688689690691692693694695696697698699700701702703704705706707708709710711712713714715716717718719720721722723724725726727728729730731732733734735736737738739740741742743744745746747748749750751752753754755756757758759760761762763764765766767768769770771772773774775776777778779780781782783784785786787788789790791792793794795796797798799800801802803804805806807808809810811812813814815816817818819820821822823824825826827828829830831832833834835836837838839840841842843844845846847848849850851852853854855856857858859860861862863864865866867868869870871872873874875876877878879880881882883884885886887888889890891892893894895896897898899900901902903904905906907908909910911912913914915916917918919920921922923924925926927928929930931932933934935936937938939940941942943944945946947948949950951952953954955956957958959960961962963964965966967968969970971972973974975976977978979980981982983984985986987988989990991992993994995996997998999100010011002100310041005100610071008100910101011101210131014101510161017101810191020102110221023102410251026102710281029103010311032103310341035103610371038103910401041104210431044104510461047104810491050105110521053105410551056105710581059106010611062106310641065106610671068106910701071107210731074107510761077107810791080108110821083108410851086108710881089109010911092109310941095109610971098109911001101110211031104110511061107110811091110111111121113111411151116111711181119112011211122112311241125112611271128112911301131113211331134113511361137113811391140114111421143114411451146114711481149115011511152115311541155115611571158115911601161116211631164116511661167116811691170117111721173117411751176117711781179118011811182118311841185118611871188118911901191119211931194119511961197119811991200120112021203120412051206120712081209121012111212121312141215121612171218121912201221122212231224122512261227122812291230123112321233123412351236123712381239124012411242124312441245124612471248124912501251125212531254125512561257125812591260126112621263126412651266126712681269127012711272127312741275127612771278127912801281128212831284128512861287128812891290129112921293129412951296129712981299130013011302130313041305130613071308130913101311131213131314

36GTCCCGAGCAACAGCTGCTGCTGCTGAGAGTCCCATGTTAAATTCAGAGCACTC 120

AGTCAAGCAGTCAAGGTGGAGGTGCTGCTAAATTTGGCTGCATCCTTTTCTCACAGCA 729

AGTCACGACAGTCAGGTGGAGGTTGCTGCTAAATTTGGCCTGCATCCTTTCTCAGACGA 180

TCGAATTTTGGCAATCTGAACCCCAAGTGAAAAAACAAATTTGCCCTGAATTGTACTGTATGTA 789

CGAATTTCATCTGAACCAAGTGAACCAAAATTCCTGATTGTACTGTATGTA 240

CTTGACACTACACAGATTCTTACCGTCTCCACAAAGGTCAGAGATTGTAATGTGCATA 849

CTGGCACTACACAGATTCTTACCGTCTCCACCAAGGTCAGAGATTGTAATGGTCAATA 300

TGACITTTTTTTTATTCCTTGACTCAGACAGCTAACTTCATTTTCAGAACTGTTTA 909

TTGACCTTTTATTCCTTGACCTCAAGACAGCTAACTTCATTTTCAGAACTGTTTA 360

ACCTTTGTGCTGTTTATAAAATATGTGTATATCCTTGTGCTTTCGTATACCA 969

Mon Nov 18 08:42:02 2002

us-09-820-003a-1.rst

Page 13

Db 361 AACCTTGTGTGCTGTTTATATAAATGATGTAATCCTTGTGCTTCCGTATACCA 420  
Qy 970 GACGTGTTCCCGTGGTGGTATATATATTTGTTGA--TGTATATATGGCATGTT 1027  
Db 421 GACTGTTCCCGTGGTGGTATATATATTTGTTGAATGTTTACTGTCATGTT 480  
Qy 1028 TAGATGTCAGTTTATGTTTGAAGATGAAGTCCACATTTGTATCAACACACAA 1087  
Db 481 TAGATGTCAGTTTATGTTTGAAGATGAAGTCCACATTTGTATCAACACACAA 540  
Qy 1088 GCAGTGTGTGCACTTCCATGATGAAGTTAGTATGATGTAAGTCTGATT 1147  
Db 541 GCAGTGTGTGCACTTCCATGATGAAGTTAGTATGATGTAAGTCTGATT 600  
Qy 1148 TGCTAGTTCTTCTGTAGAGTTATTAATGAA--AGATTACATCTGATTAATAGTTT 1206  
Db 601 TGCTAGTTCTTCTGTAGAGTTATTAATGAA--AGATTACATCTGATTAATAGTTT 660  
Qy 1207 CTTGATCTCTGATTAATGTTGGTGCAGATATGTAATTTGTCACACATATGTA 1266  
Db 661 CTTGATCTCTGATTAATGTTGGTGCAGATATGTAATTTGTCACACATATGTA 720  
Qy 1267 ACAAACAAGTGAAG 1281  
Db 721 ACAAACAAGTGAAG 735

Search completed: November 17, 2002, 13:01:48  
Job time : 1946 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 11:38:17 ; Search time 66 Seconds  
(Without alignments) 349.278 Million cell updates/sec

Title: US-09-820-003a-2

Perfect score: 886  
Sequence: 1 MSSMPEYDYLFKILLIGDS.....EKSVMKIOSTPVKQSGGCC 173

Scoring table:  
BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: A\_Geneseq.101002.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	860	97.1	205	21	AA34843
2	860	97.1	205	21	AA34844
3	817	97.1	198	21	AA34816
4	768.5	86.7	221	22	AA013525
5	768.5	86.7	224	21	AA58758
6	768.5	86.5	201	20	AA100919
7	768.5	86.5	201	22	AA28024
8	683	77.1	205	22	AA59808
9	637	71.9	202	21	AA507763
10	633	71.4	202	21	AA610858

11	619.5	69.9	258	21	AA30498	Arabidopsis thalia
12	618.5	69.8	203	21	AA30499	Arabidopsis thalia
13	612	69.1	218	21	AA30710	Arabidopsis thalia
14	612	69.1	221	21	AA30709	Arabidopsis thalia
15	612	69.1	254	21	AA45323	Arabidopsis thalia
16	580	65.5	197	21	AA48820	Arabidopsis thalia
17	544	50.1	218	21	AA19220	Arabidopsis thalia
18	442	49.9	216	21	AA08006	Arabidopsis thalia
19	431	48.6	215	21	AA35215	Zea mays protein f
20	429.5	48.5	224	21	AA47826	Arabidopsis thalia
21	429.5	48.5	234	21	AA47825	Arabidopsis thalia
22	428.5	48.4	216	21	AA08088	Arabidopsis thalia
23	428.5	48.4	253	21	AA08087	Arabidopsis thalia
24	428.5	48.4	254	21	AA08086	Arabidopsis thalia
25	428.5	48.4	254	21	AA08087	Arabidopsis thalia
26	419	47.3	213	22	AB11916	Human rab1 homolog
27	417	46.7	221	23	AB11333	Human ovary anti
28	414	46.7	207	22	AA67154	Human acid sequen
29	414	46.7	207	22	AA67154	Human acid sequen
30	409.5	46.2	335	22	AB23366	Human protein sequ
31	408	46.0	207	22	AB23366	Human protein sequ
32	405	45.7	163	21	AA18482	Drosophila melanog
33	403.5	45.5	200	21	AA18482	Drosophila melanog
34	403.5	45.5	200	21	AA18482	Drosophila melanog
35	403.5	45.5	200	21	AA18482	Drosophila melanog
36	403.5	45.5	200	21	AA18482	Drosophila melanog
37	403	45.1	141	21	AA35036	Human Rab10 protei
38	399.5	45.1	141	21	AA35036	Human Rab10 protei
39	399.5	45.1	141	21	AA35036	Human Rab10 protei
40	393	44.4	201	22	AA09982	Novel human diapo
41	393	44.4	204	22	AB70670	Drosophila melanog
42	390.5	44.1	209	22	AB23365	Novel human diapo
43	390.5	44.1	246	21	AA58196	Lung cancer associ
44	387	43.7	199	21	AA09980	Human Rab10 protei
45	387	43.7	199	21	AA09981	Human Rab10 protei

#### ALIGNMENTS

RESULT 1  
ID AAB34843 standard; Protein: 205 AA.  
AC AAB34843;  
DT 26-JAN-2001 (first entry)  
DE Gene 44 human secreted protein homologous amino acid sequence #131.  
KW Human; secreted protein; diagnosis; cytosolic; immunosuppressive;  
KW noctropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiac; gene therapy;  
KW cancer; immune disorder; cardiovascular disorder; wound healing;  
KW neurological disease; infectious disease; chromosome identification.  
OS Homo sapiens.  
PN W0200058356-A1.  
PD 05-OCT-2000.  
PE 22-MAR-2000; 2000WO-US07335.  
PR 26-MAR-1999; 99US-0126511.  
PR 17-DEC-1999; 99US-0172413.  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX WPI: 2000-594639/56.  
DR

XX Fifty nucleic acid molecules encoding human secreted proteins, useful  
PT in the prevention, treatment and diagnosis of cancer, immune disorders,  
PT cardiovascular disorders and neurological diseases -  
XX  
PS Disclosure; Page 415-416; 425pp; English.  
XX  
CC The polynucleotide sequences given in AAC59966 to AAC60015 encode the  
CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to  
CC AAB34852 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC cytoskeletal; immunosuppressive; neurotrophic; neuroprotective; antiviral;  
CC antiallergic; hepatotropic; antidiabetic; antifungal; antiparasitic; antitumor;  
CC antiviral; anticonvulsant; antibacterial; antifungal; antiparasitic; and  
CC cardiant. The polynucleotides and polypeptides are useful for  
CC preventing, treating or ameliorating a medical condition in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes for  
CC diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the gene  
CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,  
CC agonists and antagonists from the present invention are useful in the  
CC diagnosis, treatment and prevention of cancer, immune disorders,  
CC cardiovascular disorders, wound healing, neurological diseases and  
CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence  
CC used in the exemplification of the present invention.  
XX  
SQ Sequence 205 AA;  
Query Match 97.1%; Score 860; DB 21; Length 205;  
Best Local Similarity 84.4%; Pred. No. 7.6e-79;  
Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;  
QY 1 MSSNMPEYDYLFLKLLIDSGVSKSCLLRADPYTTSYSTIGVDFKRTIELDGKTI 60  
DB 1 MSSNMPEYDYLFLKLLIDSGVSKSCLLRADPYTTSYSTIGVDFKRTIELDGKTI 60  
QY 61 KLOI-----ESFNNVKWLOEIDRYASENVKL 88  
DB 61 KLOIWDTAGGERFRTITSSYRGAGLIVVYDQESFNNVKWLOEIDRYASENVKL 120  
QY 89 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNAITNEOSFMTMAEIKRKGPGA 148  
DB 121 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNAITNEOSFMTMAEIKRKGPGA 180  
QY 149 TAGGAESKNVKIQTSTPVKQSGGCC 173  
DB 181 TAGGAESKNVKIQTSTPVKQSGGCC 205  
RESULT 2  
AAB34844  
ID AAB34844 standard; Protein: 205 AA.  
XX  
AC AAB34844;  
XX  
DT 26-JAN-2001 (first entry)  
XX  
DE Human secreted protein sequence encoded by gene 44 SEQ ID NO:132.  
XX  
KW Human; secreted protein; diagnosis; cytoskeletal; immunosuppressive;  
KW neurotrophic; neuroprotective; antiviral; antiallergic; hepatotropic;  
KW antidiabetic; antineoplastic; antitumor; antiviral; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;  
KW cancer; immune disorder; cardiovascular disorder; wound healing;  
KW neurological disease; infectious disease; chromosome identification.  
XX  
OS Homo sapiens.  
XX

PN WO200058356-A1.  
XX  
XX 05-OCT-2000.  
XX  
XX 22-MAR-2000; 2000WO-US07535.  
XX  
XX 26-MAR-1999; 99US-0126511.  
XX 17-DEC-1999; 99US-0172413.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
PI Rosen CA, Ruben SM, Komatsu S G;  
XX  
XX WPI; 2000-594639/56.  
DR  
XX  
PT Fifty nucleic acid molecules encoding human secreted proteins, useful  
PT in the prevention, treatment and diagnosis of cancer, immune disorders,  
PT cardiovascular disorders and neurological diseases -  
XX  
PS Disclosure; Page 416-417; 425pp; English.  
XX  
CC The polynucleotide sequences given in AAC59966 to AAC60015 encode the  
CC human secreted proteins given in AAB34773 to AAB34822. AAB34823 to  
CC AAB34852 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC cytoskeletal; immunosuppressive; neurotrophic; neuroprotective; antiviral;  
CC antiallergic; hepatotropic; antidiabetic; antifungal; antiparasitic; antitumor;  
CC antiviral; anticonvulsant; antibacterial; antifungal; antiparasitic; and  
CC cardiant. The polynucleotides and polypeptides are useful for  
CC preventing, treating or ameliorating a medical condition in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes for  
CC diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the gene  
CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,  
CC agonists and antagonists from the present invention are useful in the  
CC diagnosis, treatment and prevention of cancer, immune disorders,  
CC cardiovascular disorders, wound healing, neurological diseases and  
CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence  
CC used in the exemplification of the present invention.  
XX  
SQ Sequence 205 AA;  
Query Match 97.1%; Score 860; DB 21; Length 205;  
Best Local Similarity 84.4%; Pred. No. 7.6e-79;  
Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;  
QY 1 MSSNMPEYDYLFLKLLIDSGVSKSCLLRADPYTTSYSTIGVDFKRTIELDGKTI 60  
DB 1 MSSNMPEYDYLFLKLLIDSGVSKSCLLRADPYTTSYSTIGVDFKRTIELDGKTI 60  
QY 61 KLOI-----ESFNNVKWLOEIDRYASENVKL 88  
DB 61 KLOIWDTAGGERFRTITSSYRGAGLIVVYDQESFNNVKWLOEIDRYASENVKL 120  
QY 89 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNAITNEOSFMTMAEIKRKGPGA 148  
DB 121 LVGNKCDLTTKKVVDYTTAKFADSLGIPLETSAKNAITNEOSFMTMAEIKRKGPGA 180  
QY 149 TAGGAESKNVKIQTSTPVKQSGGCC 173  
DB 181 TAGGAESKNVKIQTSTPVKQSGGCC 205  
RESULT 3  
AAB34816  
ID AAB34816 standard; Protein: 198 AA.  
XX  
AC AAB34816;  
XX

XX 26-JAN-2001 (first entry)  
DT Human secreted protein sequence encoded by gene 44 SEQ ID NO:104.  
XX  
DE Human secreted protein sequence encoded by gene 44 SEQ ID NO:104.  
DE  
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;  
KW neotropic; neuroprotective; antiviral; antidiabetic; hepatotropic;  
KW antidiabetic; antineoplastic; antitumor; antiparasitic; anticonvulsant;  
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;  
KW cancer; immune disorder; cardiovascular disorder; wound healing;  
KW neurological disease; infectious disease; chromosome identification.  
XX  
OS Homo sapiens.  
XX  
PN WO200058356-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 22-MAR-2000; 2000MO-US07535.  
XX  
PR 26-MAR-1999; 99US-0126511.  
PR 17-DEC-1999; 99US-0172413.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM, Komatsoulis G;  
XX  
DR WPI; 2000-594639/56.  
DR N-PSDB; AAC60009.  
XX  
PT Fifty nucleic acid molecules encoding human secreted proteins, useful  
PT in the prevention, treatment and diagnosis of cancer, immune disorders,  
PT cardiovascular disorders and neurological diseases -  
XX  
PS Claim 1; Page 393-394; 425pp; English.  
XX  
XX The polynucleotide sequences given in AAC59966 to AAC60015 encode the  
CC human secreted proteins given in AAB34773 to AAB34823 to  
CC AAB34852 represent human secreted polypeptide sequences and proteins  
CC homologous to them, which are given in the exemplification of the present  
CC invention. Human secreted proteins have activities based on the tissues  
CC and cells the genes are expressed in. Examples of activities include:  
CC cytostatic; immunosuppressive; neotropic; neuroprotective; antiviral;  
CC antidiabetic; hepatotropic; antidiabetic; antineoplastic; antitumor;  
CC vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic; and  
CC cardiant. The polynucleotides and polypeptides are useful for  
CC preventing, treating or ameliorating a medical condition in e.g. humans,  
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The  
CC polypeptides can also be used as a food additive or preservative to  
CC increase or decrease storage capabilities. The polynucleotides are  
CC useful for chromosome identification. They are also useful as probes for  
CC diagnosing a disorder related to the female reproductive system,  
CC particularly breast and/or ovary cancer. They are also useful in the gene  
CC therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,  
CC agonists and antagonists from the present invention are useful in the  
CC diagnosis, treatment and prevention of cancer, immune disorders and  
CC cardiovascular disorders, wound healing, neurological diseases and  
CC infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence  
CC used in the exemplification of the present invention.  
XX  
SQ Sequence 198 AA:  
XX  
Query Match 92.2%; Score 817; DB 21; Length 198;  
Best Local Similarity 83.8%; Pred. No. 1.6e-74;  
Matches 165; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

OY 97 TTKKVDYTTAKFADSLGIPFLETSKAKNATNVEOSFMTMAEIKRMPGATAGAKRS 156  
DB 121 TTKKVDYTTAKFADSLGIPFLETSKAKNATNVEOSFMTMAEIKRMPGATAGAKRS 180  
OY 157 NVKIOSPPVKOSGGGCC 173  
DB 181 NVKIOSPPVKOSGGGCC 197  
RESULT 4  
ID AAO13525 standard; Protein; 221 AA.  
XX AAO13525  
XX AAO13525;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 27417.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-514838/56.  
DR N-PSDB; AA193456.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 27417; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 221 AA:  
XX  
Query Match 86.7%; Score 768.5; DB 22; Length 221;  
Best Local Similarity 76.5%; Pred. No. 1.5e-69;  
Matches 156; Conservative 8; Mismatches 7; Indels 33; Gaps 2;

OY 2 SSMPEDYDLFTLLIGDSGVGKSCLLRFADDTYTSYSTGTGVDFKRIETEDGKTIK 61  
DB 19 AANPEYDLFTLLIGDSGVGKSCLLRFADDTYTSYSTGTGVDFKRIETEDGKTIK 78

QY 62 LQI-----ESENWYKQWLOEIDRYASENVKLL 89  
 Db 79 LQIWDTAGQERFRITSSYRGAGHIIYVDYDTQESYANKWLOEIDRYASENVKLL 138  
 QY 90 VGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNVQSFMTMAEIKRKGPGAT 149  
 Db 139 VGNKSDLTTRKVVNDYTTAKAFADSLGIPLETSAKNATNVQSFMTMAEIKRKGPGAA 198  
 QY 150 AGAEKSNVRIQSTPVPKQSGGCC 173  
 Db 199 SGG-ERPLKIDSTPVPKAGGCC 221

RESULT 5  
 AAB58758  
 ID AAB58758 standard; Protein: 224 AA.  
 AC AAB58758;  
 XX  
 D7 27-MAR-2001 (first entry)  
 XX  
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 466.  
 XX  
 KM Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
 KM nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
 KM antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;  
 KM antibacterial; antifungal; antiparasitic; cardiact; immune disorder;  
 KM Addison's disease; allergy; autoimmune haemolytic anaemia;  
 KM autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KM multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KM cardiovascular disorder; wound healing; neurological disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200055173-A1.  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05881.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI: 2000-611515/58.  
 DR N-PADB; AAF21661.  
 XX  
 XX New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention,  
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
 PT disorders and neurological diseases -  
 XX  
 PS Claim 11; Page 895-896; 1299pp; English.  
 XX  
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the  
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
 CC nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
 CC antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;  
 CC antibacterial; antifungal; antiparasitic and cardiact activity. The  
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
 CC particularly breast and ovarian cancer. The nucleic acid sequences,  
 CC proteins, agonists and antagonists may also be used in the diagnosis,  
 CC prevention and treatment of immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

CC arthritis and ulcerative colitis; cardiovascular disorders such as  
 CC myocardial ischaemias; wound healing; neurological diseases such as  
 CC cerebral anoxia and epilepsy; and infectious diseases.  
 XX  
 SQ Sequence 224 AA;  
 Query Match 86.7%; Score 768.5; DB 21; Length 224;  
 Best Local Similarity 76.5%; Pred. No. 1.5e-69;  
 Matches 156; Conservative 8; Mismatches 7; Indels 33; Gaps 2;

QY 2 SSANPEYDYLKLLLLIGDSGVKSCLLRFPADPTYSYSTIVDPKIRIIEIDGKTK 61  
 Db 22 AANPEYDYLKLLLLIGDSGVKSCLLRFPADPTYSYSTIVDPKIRIIEIDGKTK 81  
 QY 62 LQI-----ESENWYKQWLOEIDRYASENVKLL 89  
 Db 82 LQIWDTAGQERFRITSSYRGAGHIIYVDYDTQESYANKWLOEIDRYASENVKLL 141  
 QY 90 VGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNVQSFMTMAEIKRKGPGAT 149  
 Db 142 VGNKSDLTTRKVVNDYTTAKAFADSLGIPLETSAKNATNVQSFMTMAEIKRKGPGAA 201  
 QY 150 AGAEKSNVRIQSTPVPKQSGGCC 173  
 Db 202 SGG-ERPLKIDSTPVPKAGGCC 224

RESULT 6  
 AAY00919  
 ID AAY00919 standard; Protein: 201 AA.  
 AC AAY00919;  
 XX  
 DT 28-MAY-1999 (first entry)  
 XX  
 DE Human Rab protein, RABP-2, protein sequence.  
 XX  
 KM Rab protein; RABP-1; RABP-2; RABP-3; human; vesicle trafficking; cancer;  
 KM cell differentiation; apoptosis; immunodeficiency; cell proliferation;  
 KM neurodegenerative disease; myelodysplastic syndrome; wasting disease;  
 KM toxin-induced disease; infection; genetic defect; diagnosis; therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W090909182-A2.  
 PD 25-FEB-1999.  
 XX  
 PF 17-AUG-1998; 98WO-US16983.  
 XX  
 PR 21-AUG-1997; 97US-0916901.  
 XX  
 PA (INCY-) INCYTE PHARM INC.  
 PI Corley NC, Hillman JL, Lal P, Shah P;  
 XX  
 DR WPI: 1999-181042/15.  
 DR N-PADB; AAX27232.  
 XX  
 PT New purified human Rab proteins - used to develop products for  
 PT treating e.g. AIDS, immunodeficiencies, neurodegenerative diseases,  
 PT cancer, inflammation or autoimmune diseases.  
 XX  
 PS Claim 23; Fig 2; 94pp; English.  
 XX  
 CC This sequence is a human Rab protein of the invention, designated  
 CC RABP-2. The Rab proteins, RABP-1, RABP-2 and RABP-3 are involved in  
 CC vesicle trafficking, cell function, and cell differentiation. The RABP  
 CC polypeptides, DNAs and agonists can be used to prevent or treat a  
 CC disorder associated with an increase in apoptosis, e.g. infectious or  
 CC genetic immunodeficiencies, neurodegenerative diseases such as  
 CC Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis,  
 CC retinitis pigmentosa, and cerebellar degeneration, myelodysplastic



CC syndromes such as aplastic anaemia, ischaemic injuries such as myocardial  
 CC infarction, stroke, and reperfusion injury, toxin-induced diseases such  
 CC as alcohol-induced liver damage, cirrhosis, and lathyrism, wasting  
 CC diseases such as cachexia, viral infections, and osteoporosis. They can  
 CC also be used to stimulate cell proliferation for use in transplantation  
 CC or to produce cells to fight an infection or a cancer or to correct a  
 CC genetic defect in a disease such as sickle cell beta thalassemia, cystic  
 CC fibrosis or Huntington's chorea. Antagonists can be used to prevent or  
 CC treat a disorder associated with cell proliferation e.g. cancers or  
 CC inflammation, e.g. Addison's disease, adult respiratory distress  
 CC syndrome, allergies, asthma, atherosclerosis, bronchitis, cholecystitis,  
 CC Crohn's disease, ulcerative colitis, atopic dermatitis, glomerulonephritis,  
 CC gout, Graves' disease, hyperosinophilia, irritable bowel syndrome, lupus  
 CC erythematosus, multiple sclerosis, myasthenia gravis, inflammation,  
 CC osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid  
 CC arthritis, scleroderma, Sjogren's syndrome, and autoimmune thyroiditis,  
 CC complications of cancer, haemodialysis, extracorporeal circulation,  
 CC infections and trauma.

XX Sequence 201 AA:  
 SQ

Query Match 86.5%; Score 766.5; DB 20; Length 201;  
 Best local similarity 77.2%; Pred. No. 2,1e-69;  
 Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

QY 4 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESYSTGVPRKRTETEDGKTIKQ 63  
 DB 1 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESYSTGVPRKRTETEDGKTIKQ 60  
 QY 64 I-----ESPNNVKOMLOEDIRYASENVKLLYG 91  
 DB 61 IMDTGAQERFRTITSSYRGAAGIIYVDVTDQESYANKOMLOEDIRYASEVKNKLLYG 120  
 QY 92 NKCDLTTKKVVDYTTAKEPADSLGIPLETSAKNATNVDSFTMAAEIKRRMGPGATAG 151  
 DB 121 NKSDLTTKKVVDTTAKFPADSLGIPLETSAKNATNVDSFTMAAEIKRRMGPGASG 180  
 QY 152 GAESNVKIOSTPVKOSGGCC 173  
 DB 181 G-ERPNIKIDSTVPKAGGCC 201

RESULT 7  
 AAU28024  
 ID AAU28024 standard; Protein: 201 AA.  
 AC AAU28024:  
 XX  
 XX  
 DT 18-DEC-2001 (first entry)  
 DE Novel human secretory protein, Seq ID No 193.  
 XX  
 KM Human: secreted protein; arthritis; Crohn's disease; sepsis; shock;  
 KM ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
 KM transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
 KM amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
 KM ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
 KM gut protection; lung; liver fibrosis; immune deficiency; infection;  
 KM severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
 KM multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
 KM fertility; analgesic; pain; antigen.  
 XX  
 OS Homo sapiens.  
 PN MO200166689-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 05-MAR-2001; 2001MO-US04942.  
 XX  
 PR 07-MAR-2000; 2000US-0519705.  
 PR 19-MAY-2000; 2000US-0574434.

PR 17-JUN-2000; 2000US-0596193.  
 PR 14-JUL-2000; 2000US-0616847.  
 PR 19-SEP-2000; 2000US-0665363.  
 PR 20-OCT-2000; 2000US-0693267.  
 XX  
 PA (HYSE-) HYSEO INC.  
 PI Tang Y<sup>1</sup>, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;  
 PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;  
 DR N-PSDB: AAS44924.  
 DR WPI: 2001-589934/66.  
 PS Novel polypeptides and nucleic acids obtained from cDNA libraries  
 PT prepared from various human tissues, for diagnosis and treatment of  
 PT cancer, neurological, inflammatory, and autoimmune disorders -  
 CC Example 3; Seq ID No 193; 107pp; English.

XX The invention relates to novel isolated human secreted polypeptides (I)  
 CC and polynucleotides (II). (I) and (II) are useful for treating  
 CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,  
 CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is  
 CC involved in increasing haematopoiesis, stem cell survival, bone growth  
 CC and remodeling. (I), (II) and modulators of (II) are useful for  
 CC prophylaxis or treatment of one or more cancers. (II) is also useful for  
 CC creating transgenic animals useful for studying the in vivo activities of  
 CC the polypeptide as well as for studying modulators of the polypeptides.  
 CC (I) induces the proliferation of neural cells and regeneration of nerve  
 CC and brain tissue and is useful for the treatment of central and  
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
 CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis and is useful for treating myeloid  
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
 CC tissue growth, and in tissue repair, healing of burns, incisions,  
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
 CC disorders, or periodontal disease. Furthermore, (I) is also useful for  
 CC gut protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, various immune deficiencies and  
 CC disorders including severe combined immunodeficiency (SCID), bacterial or  
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
 CC reactions and conditions, such as asthma or other respiratory problems.  
 CC In addition, (I) affects biorhythms or circadian cycles of rhythms,  
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
 CC analgesic effects or other pain reducing effects, immunoglobulin like  
 CC activity and can act as an antigen in a vaccine composition to raise an  
 CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
 CC amino acid sequences of the invention.

XX Sequence 201 AA:  
 SQ

Query Match 86.5%; Score 766.5; DB 22; Length 201;  
 Best local similarity 77.2%; Pred. No. 2,1e-69;  
 Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

QY 4 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESYSTGVPRKRTETEDGKTIKQ 63  
 DB 1 MNPEVDYLFKLLIGDSGVKSCLLRFADDTYESYSTGVPRKRTETEDGKTIKQ 60  
 QY 64 I-----ESPNNVKOMLOEDIRYASENVKLLYG 91  
 DB 61 IMDTGAQERFRTITSSYRGAAGIIYVDVTDQESYANKOMLOEDIRYASEVKNKLLYG 120  
 QY 92 NKCDLTTKKVVDYTTAKEPADSLGIPLETSAKNATNVDSFTMAAEIKRRMGPGATAG 151  
 DB 121 NKSDLTTKKVVDTTAKFPADSLGIPLETSAKNATNVDSFTMAAEIKRRMGPGASG 180  
 QY 152 GAESNVKIOSTPVKOSGGCC 173  
 DB 181 G-ERPNIKIDSTVPKAGGCC 201

Db 181 G-ERPNTKIDSTPVKPGAGGCC 201

RESULT 8

ABB59808 ID ABB59808 standard; Protein: 205 AA.

XX ABB59808;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 6216.

KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

XX W0200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001MO-US09231.

PR 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li FWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL03911.

PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Disclosure; SEQ ID NO 6216; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences and pharmaceutical drugs. The invention CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 205 AA;

Query Match 77.1%; Score 683; DB 22; Length 205;

Best Local Similarity 69.4%; Pred. No. 6e-61; Matches 143; Conservative 12; Mismatches 17; Indels 34; Gaps 3;

QY 1 MSSNNPEVDYLFKLLIGDSGVKSCLLRRADTYTTSYSTIGVDKIRTIELDGKTI 60

Db 1 MSSVNPEDYLFKLLIGDSGVKSCLLRRADTYTTSYSTIGVDKIRTIELDGKTI 60

QY 61 KLQI-----ESFNWVKWLOEIDRYASENVKL 88

Db 61 KLQIWDTAGOERFRTITSSYRGAGIIVVDCIDQESFNWVKWLOEIERACENVKL 120

QY 89 LVGNKCDLTTKRVYDYTTAKFADSLGIPLETSAKNATNVEQSFMTAAEIKRMGPGA 148

Db 121 LVGNKSDLTTRKRVYDHTTAAYAAQIGIPLETSAKSAKTNVEQAEMTAAEIKRNVGPS 180

QY 149 TAGGAEKSNVAI-OSTPVKQSGGCC 173

Db 181 SATD-NASKVKIDQGRPVENKSGCC 205

RESULT 9

AA007763 ID AA007763 standard; Protein: 202 AA.

XX AA007763;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 5040.

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126254.

XX 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

XX 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

XX 17-JUN-1999; 99US-0139452.

PR 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142390.  
PR 06-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144862.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147499.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148568.  
PR 16-AUG-1999; 99US-0148598.  
PR 17-AUG-1999; 99US-0148175.  
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158023.  
PR 08-OCT-1999; 99US-0158322.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 71.9%; Score 637; DB 21; Length 202;  
Best Local Similarity 65.3%; Pred. No 2 6e-56;  
Matches 132; Conservative 11; Mismatches 25; Indels 34; Gaps 3;  
  
QY 4 MNEBYVYKLLIGDSGCKSCLLRPADDTYESTISITGVNDEKRTITLCKTIKQ 63  
DB 1 MNEBYDLKLLIGDSGVKSCLLKRPADSDYDSTISITGVNDEKRTITVQDKTIKQ 60  
QY 64 I-----ESPNNYKQWLOEIDRYASEVNNKLLYG 91  
DB 61 IWPDAQERFTITSSYRKAGHGIYTVYDVTDLESFPNNYKQWLMNEIDRYASEVNNKLLYG 120  
QY 92 NKCDLTKKVVDTYTAKEPADSIGIPLETISAKNATNVEQSFMTAAEIKRRMGPGATAG 151  
|||||:|||| ||| ||| |||||||||||||||:|||| |||



```
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149476.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158028.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158309.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159684.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          71.4%  Score 633.  DB 21:  Length 202.
Best Local Similarity 65.8%  Pred No 6 7e-56.
Matches 133;  Conservative 10;  Mismatches 25;  Indels 34;  Gaps 3.
```

```
DB 61 1WPTAQEFERTITSSYYRGAGSIIYVDYDLESEFNNVKOMLNIDIVYSEVNNKLLVG 120
OY 92 NKCDLITKKVDDYTTAKEFADSLGIFLETSAKNATNVEQSFMTAAEIKKRGPGATAG 151
DB 121 NKNDLISOKVSTEFARAFADDELGIPILETSAKNATNVEEAFMAMTAIKTRNA-SOPAG 179
OY 152 GAESKNVKIOSTPYKQSGGGCC 173
DB 180 GAKPPTVQIRGQPVNOC-SGCC 200

RESULT 11
AAG30498
ID AAG30498 standard; Protein; 258 AA.
AC AAG30498;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 36472.
KW Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132408.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134421.
PR 14-MAY-1999; 99US-0134470.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137722.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
```

PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140685.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142350.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145224.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 69.94; Score 619.5; DB 21; Length 258;  
Best Local Similarity 61.64; Pred. No. 2.1e-54;  
Matches 125; Conservative 15; Mismatches 30; Indels 33; Gaps 2;

```
OY 3 SMMPEYDLFLKLLIDSGVSKCLLLRFADDPYTESYISTIGVDFKTRTELIDCKITKL 62
Db 55 TAMPEYDLFLKLLIDSGVSKCLLLRFADDPYTESYISTIGVDFKTRTELIDCKITKL 114
OY 63 QI-----EFNNYKQMLQEDIRASNNKLLV 90
Db 115 QIMDPAQGERRITTSYRGAGIIVDYTBESFNNAKMLSEIDIRASDNNKLLV 174
OY 91 GKKCLTITKKKYVDYTTAKEPADSLGIPPLETSKKNATVNEOSFMTMAETKKRMPGATA 150
Db 175 GKNKSLTENRAIPETAKAFADDELGIPPEMETSADKATVNEQAFMAASASIKERMAQAPAG 234
OY 151 GGAESKNVKGISTPVPKOSGGCC 173
Db 235 NNARPTVOIRGQPVAKK-NGCC 256

RESULT 12
AAG30499
ID AAG30499 standard; Protein; 203 AA.
AC AAG30499;
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 36473.
XX
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
PN
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PE
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0122548.
PR 22-MAR-1999; 99US-0125789.
PR 22-MAR-1999; 99US-0126284.
PR 23-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0128233.
PR 06-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141824.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 26-JUL-1999; 99US-0145526.
PR 26-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145518.
PR 27-JUL-1999; 99US-0145519.
PR 28-JUL-1999; 99US-0145581.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
```

PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151060.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154079.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 14-OCT-1999; 99US-0159584.  
PR 18-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160960.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 69.8%; Score 618.5; DB 21; Length 203;  
Best Local Similarity 61.9%; Pred. No. 2e-54;  
Matches 125; Conservative 14; Mismatches 30; Indels 33; Gaps 2;  
  
QY 4 MNPEDYLFKLLIGDSVGSCKLLRPADDTYESYSTGVDPKRTIILDGKTIKQ 63  
Db 1 MNPEDYLFKLLIGDSVGSCKLLRPADDTYESYSTGVDPKRTIILDGKTIKQ 60  
QY 64 I-----ESFNNVKOMLOEDIRYASENFKILV 91  
Db 61 IWDTAGOERFRTITSSYRGAGHIIIVDYDESFNNVKOMLSEIDRTASDNVKKLIV 120  
QY 92 NKCDLTKRVYDTTAKFADSLGIPLETSAKNATNVEQSFMTAAEIKRMGPATAG 151  
Db 121 NKSDLTENRAIPYETAKAFADIEIGIPFETSAKDATNVEQAFMMSASIKERMASQPAQN 180  
QY 152 GAESKNVKIOSTPVKQSGGCC 173  
Db 181 NARPTVOIRGQPYAOK-NGCC 201  
  
RESULT 13  
AAG30710  
ID AAG30710 standard; Protein; 218 AA.  
XX AAG30710;  
AC  
XX  
XX 17-OCT-2000 (first entry)  
DE  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 36764.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX Arabidopsis thaliana.  
OS  
XX EPI033405-A2.  
PN  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.



Mon Nov 18 08:42:02 2002

us-09-820-003a-2.rag

Page 13

PR	20-MAY-1999;	9905-0135124.	PR	02-AUG-1999;	9905-0146388.
PR	21-MAY-1999;	9905-0135353.	PR	02-AUG-1999;	9905-0146389.
PR	24-MAY-1999;	9905-0135629.	PR	03-AUG-1999;	9905-0147038.
PR	25-MAY-1999;	9905-0136021.	PR	04-AUG-1999;	9905-0147204.
PR	27-MAY-1999;	9905-0136392.	PR	04-AUG-1999;	9905-0147302.
PR	28-MAY-1999;	9905-0136782.	PR	05-AUG-1999;	9905-0147192.
PR	01-JUN-1999;	9905-0137222.	PR	05-AUG-1999;	9905-0147260.
PR	03-JUN-1999;	9905-0137528.	PR	06-AUG-1999;	9905-0147303.
PR	04-JUN-1999;	9905-0137502.	PR	06-AUG-1999;	9905-0147416.
PR	07-JUN-1999;	9905-0137724.	PR	09-AUG-1999;	9905-0147493.
PR	08-JUN-1999;	9905-0138094.	PR	09-AUG-1999;	9905-0147935.
PR	10-JUN-1999;	9905-0138540.	PR	10-AUG-1999;	9905-0148171.
PR	10-JUN-1999;	9905-0138847.	PR	11-AUG-1999;	9905-0148319.
PR	14-JUN-1999;	9905-0139119.	PR	12-AUG-1999;	9905-0148341.
PR	16-JUN-1999;	9905-0139452.	PR	13-AUG-1999;	9905-0148565.
PR	16-JUN-1999;	9905-0139453.	PR	13-AUG-1999;	9905-0149368.
PR	17-JUN-1999;	9905-0139492.	PR	16-AUG-1999;	9905-0149175.
PR	18-JUN-1999;	9905-0139454.	PR	17-AUG-1999;	9905-0149175.
PR	18-JUN-1999;	9905-0139455.	PR	18-AUG-1999;	9905-0149426.
PR	18-JUN-1999;	9905-0139456.	PR	20-AUG-1999;	9905-0149722.
PR	18-JUN-1999;	9905-0139457.	PR	20-AUG-1999;	9905-0149723.
PR	18-JUN-1999;	9905-0139458.	PR	20-AUG-1999;	9905-0149929.
PR	18-JUN-1999;	9905-0139459.	PR	23-AUG-1999;	9905-0149930.
PR	18-JUN-1999;	9905-0139460.	PR	23-AUG-1999;	9905-0149930.
PR	18-JUN-1999;	9905-0139461.	PR	25-AUG-1999;	9905-0150566.
PR	18-JUN-1999;	9905-0139462.	PR	26-AUG-1999;	9905-0150884.
PR	18-JUN-1999;	9905-0139463.	PR	27-AUG-1999;	9905-0151065.
PR	18-JUN-1999;	9905-0139464.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139465.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139466.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139467.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139468.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139469.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139470.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139471.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139472.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139473.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139474.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139475.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139476.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139477.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139478.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139479.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139480.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139481.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139482.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139483.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139484.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139485.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139486.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139487.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139488.	PR	27-AUG-1999;	9905-0151066.
PR	18-JUN-1999;	9905-0139489.			

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 69.18; Score 612; DB 21; Length 218;  
Best Local Similarity 65.08; Pred. No. 9.8e-54;  
Matches 128; Conservative 11; Mismatches 24; Indels 34; Gaps 3;

QY 9 DYLFKLLIGSGVSKSCLLRFADDDTSESYSISIGVDFKRTIELDCKTIIQI---- 64  
DB 22 DYLFKLLIGSGVSKSCLLRFADDDTSESYSISIGVDFKRTIELDCKTIIQIWPDA 81  
QY 65 -----ESFNNVKOMLOEIDRYASENVNKLIVGNKCDL 96  
DB 82 GQERFRTTSSYRGAGHIIYTDVTDLESFNNVKOMLEIDRYASENVNKLIVGNKCDL 141  
QY 97 TTKRVVDYTTAKERPADSIGIPLETSAKATNVEQSFMTMAEIKRKGPGATAGAEKS 156  
DB 142 TSQKRVSTETAKAFADDELIGIPLETSAKATNVEEAFYMAAIAIKTRMA-SQPAQSKRP 200  
QY 157 NVKIQSPYKSGGCC 173  
DB 201 TVQIRGPVNOQ-SGCC 216

## RESULT 14

AAG30709  
ID AAG30709 standard; Protein: 221 AA.

XX AAG30709;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 36763.

DE Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142920.  
PR 13-JUL-1999; 99US-0142977.  
PR 14-JUL-1999; 99US-0143542.  
PR 15-JUL-1999; 99US-0143624.  
PR 16-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0147936.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149358.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157865.  
PR 06-OCT-1999; 99US-0158029.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 13-OCT-1999; 99US-0159325.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 69.1%; Score 612; DB 21; Length 221;  
Best Local Similarity 65.0%; Pred. No. 1e-53;  
Matches 18; Conservative 11; Mismatches 24; Indels 34; Gaps 3;  
  
QY 9 DYLFKLLIGDSGVKSCLLRPADDTYESYSTIGVDPKIRITELDGKTIKLOI---- 64  
DB 25 DYLFKLLIGDSGVKSCLLRPADDSYLDYSITIGVDPKIRITVEQDKTIKLOIMDPA 84  
  
QY 65 -----ESFNNYKOWLEIDIRYASEVYKLLVGNKKDL 96  
DB 85 GQRFRTTSSYYGAGGIVTYDVTDLSEFNNYKOWLEIDIRYASEVYKLLVGNKKDL 144  
  
QY 97 TTKKVVDTTAKERADSLGIPELETSANNATNVEQSFMTAAEIKRMGPGATAGAEKS 156  
DB 145 TSQKVVSTETAKAFADSLGIPELETSANNATNVEAFMTAIXTRMA-SQPGASRPP 203  
  
QY 157 NVRIGSTPYKQSGGCC 173  
DB 204 TVQIRQPFVNOQ-SGCC 219  
  
RESUME 15  
AKG45323 standard; Protein; 254 AA.  
ID AKG45323  
AC AAG45323;  
XX  
XX  
DP 18-OCT-2000 (first entry)  
DE  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 56887.  
XX  
XX Protein identification: signal transduction pathway; metabolic pathway;  
KW hybridization assay; genetic mapping; gene expression control; promoter;  
KM termination sequence.  
XX  
OS Arabidopsis thaliana.  
PN  
PN EPI033405-A2.  
PD  
PD 06-SEP-2000.  
XX  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 9905-0132486.  
 PR 06-MAY-1999; 9905-0132487.  
 PR 07-MAY-1999; 9905-0132863.  
 PR 11-MAY-1999; 9905-0134256.  
 PR 14-MAY-1999; 9905-0134218.  
 PR 14-MAY-1999; 9905-0134219.  
 PR 14-MAY-1999; 9905-0134221.  
 PR 14-MAY-1999; 9905-0134370.  
 PR 18-MAY-1999; 9905-0134768.  
 PR 19-MAY-1999; 9905-0134941.  
 PR 20-MAY-1999; 9905-0135124.  
 PR 21-MAY-1999; 9905-0135353.  
 PR 24-MAY-1999; 9905-0135629.  
 PR 25-MAY-1999; 9905-0136021.  
 PR 27-MAY-1999; 9905-0136392.  
 PR 28-MAY-1999; 9905-0136782.  
 PR 01-JUN-1999; 9905-0137222.  
 PR 03-JUN-1999; 9905-0137528.  
 PR 04-JUN-1999; 9905-0137502.  
 PR 07-JUN-1999; 9905-0137724.  
 PR 08-JUN-1999; 9905-0138094.  
 PR 10-JUN-1999; 9905-0138540.  
 PR 10-JUN-1999; 9905-0138847.  
 PR 14-JUN-1999; 9905-0139119.  
 PR 16-JUN-1999; 9905-0139432.  
 PR 17-JUN-1999; 9905-0139453.  
 PR 18-JUN-1999; 9905-0139452.  
 PR 18-JUN-1999; 9905-0139456.  
 PR 18-JUN-1999; 9905-0139457.  
 PR 18-JUN-1999; 9905-0139458.  
 PR 18-JUN-1999; 9905-0139459.  
 PR 18-JUN-1999; 9905-0139460.  
 PR 18-JUN-1999; 9905-0139461.  
 PR 18-JUN-1999; 9905-0139462.  
 PR 18-JUN-1999; 9905-0139463.  
 PR 18-JUN-1999; 9905-0139750.  
 PR 18-JUN-1999; 9905-0139763.  
 PR 21-JUN-1999; 9905-0139817.  
 PR 22-JUN-1999; 9905-0139899.  
 PR 23-JUN-1999; 9905-0140353.  
 PR 23-JUN-1999; 9905-0140354.  
 PR 24-JUN-1999; 9905-0140695.  
 PR 28-JUN-1999; 9905-0140823.  
 PR 29-JUN-1999; 9905-0140921.  
 PR 30-JUN-1999; 9905-0141287.  
 PR 01-JUL-1999; 9905-0141842.  
 PR 01-JUL-1999; 9905-0142154.  
 PR 02-JUL-1999; 9905-0142055.  
 PR 06-JUL-1999; 9905-0142390.  
 PR 08-JUL-1999; 9905-0142803.  
 PR 09-JUL-1999; 9905-0142920.  
 PR 12-JUL-1999; 9905-0142977.  
 PR 13-JUL-1999; 9905-0143542.  
 PR 14-JUL-1999; 9905-0143624.  
 PR 15-JUL-1999; 9905-0144005.  
 PR 16-JUL-1999; 9905-0144085.  
 PR 16-JUL-1999; 9905-0144086.  
 PR 19-JUL-1999; 9905-0144325.  
 PR 19-JUL-1999; 9905-0144331.  
 PR 19-JUL-1999; 9905-0144332.  
 PR 19-JUL-1999; 9905-0144333.  
 PR 19-JUL-1999; 9905-0144334.  
 PR 19-JUL-1999; 9905-0144335.  
 PR 20-JUL-1999; 9905-0144352.  
 PR 20-JUL-1999; 9905-0144632.  
 PR 20-JUL-1999; 9905-0144864.  
 PR 21-JUL-1999; 9905-0144814.  
 PR 21-JUL-1999; 9905-0145086.  
 PR 22-JUL-1999; 9905-0145088.  
 PR 22-JUL-1999; 9905-0145085.  
 PR 22-JUL-1999; 9905-0145087.

PR 22-JUL-1999; 9905-0145089.  
 PR 22-JUL-1999; 9905-0145192.  
 PR 23-JUL-1999; 9905-0145145.  
 PR 23-JUL-1999; 9905-0145218.  
 PR 23-JUL-1999; 9905-0145224.  
 PR 26-JUL-1999; 9905-0145276.  
 PR 27-JUL-1999; 9905-0145913.  
 PR 27-JUL-1999; 9905-0145918.  
 PR 27-JUL-1999; 9905-0145919.  
 PR 28-JUL-1999; 9905-0145951.  
 PR 02-AUG-1999; 9905-0146386.  
 PR 02-AUG-1999; 9905-0146388.  
 PR 02-AUG-1999; 9905-0146389.  
 PR 03-AUG-1999; 9905-0147038.  
 PR 04-AUG-1999; 9905-0147204.  
 PR 04-AUG-1999; 9905-0147302.  
 PR 05-AUG-1999; 9905-0147192.  
 PR 05-AUG-1999; 9905-0147260.  
 PR 06-AUG-1999; 9905-0147303.  
 PR 06-AUG-1999; 9905-0147416.  
 PR 09-AUG-1999; 9905-0147493.  
 PR 09-AUG-1999; 9905-0147935.  
 PR 10-AUG-1999; 9905-0148171.  
 PR 11-AUG-1999; 9905-0148319.  
 PR 12-AUG-1999; 9905-0148341.  
 PR 13-AUG-1999; 9905-0148565.  
 PR 13-AUG-1999; 9905-0148684.  
 PR 16-AUG-1999; 9905-0149368.  
 PR 17-AUG-1999; 9905-0149175.  
 PR 18-AUG-1999; 9905-0149426.  
 PR 20-AUG-1999; 9905-0149722.  
 PR 20-AUG-1999; 9905-0149723.  
 PR 20-AUG-1999; 9905-0149929.  
 PR 23-AUG-1999; 9905-0149902.  
 PR 23-AUG-1999; 9905-0149930.  
 PR 23-AUG-1999; 9905-0150566.  
 PR 26-AUG-1999; 9905-0150884.  
 PR 27-AUG-1999; 9905-0151065.  
 PR 27-AUG-1999; 9905-0151066.  
 PR 27-AUG-1999; 9905-0151080.  
 PR 30-AUG-1999; 9905-0151303.  
 PR 31-AUG-1999; 9905-0151438.  
 PR 01-SEP-1999; 9905-0151930.  
 PR 07-SEP-1999; 9905-0152363.  
 PR 10-SEP-1999; 9905-0153070.  
 PR 13-SEP-1999; 9905-0153758.  
 PR 15-SEP-1999; 9905-0154018.  
 PR 16-SEP-1999; 9905-0154039.  
 PR 20-SEP-1999; 9905-0154779.  
 PR 22-SEP-1999; 9905-0155139.  
 PR 23-SEP-1999; 9905-0155486.  
 PR 24-SEP-1999; 9905-0155659.  
 PR 26-SEP-1999; 9905-0156458.  
 PR 29-SEP-1999; 9905-0156596.  
 PR 04-OCT-1999; 9905-0157117.  
 PR 05-OCT-1999; 9905-0157753.  
 PR 06-OCT-1999; 9905-0157865.  
 PR 07-OCT-1999; 9905-0158029.  
 PR 08-OCT-1999; 9905-0158232.  
 PR 12-OCT-1999; 9905-0158369.  
 PR 13-OCT-1999; 9905-0159293.  
 PR 13-OCT-1999; 9905-0159294.  
 PR 13-OCT-1999; 9905-0159295.  
 PR 14-OCT-1999; 9905-0159329.  
 PR 14-OCT-1999; 9905-0159330.  
 PR 14-OCT-1999; 9905-0159331.  
 PR 14-OCT-1999; 9905-0159637.  
 PR 14-OCT-1999; 9905-0159638.  
 PR 18-OCT-1999; 9905-0159584.  
 PR 18-OCT-1999; 9905-0160741.  
 PR 21-OCT-1999; 9905-0160767.  
 PR 21-OCT-1999; 9905-0160768.  
 PR 21-OCT-1999; 9905-0160770.





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 13:04:12 : Search time 20 Seconds  
(without alignments)  
130.277 Million cell updates/sec

Title: US-09-820-003a-2  
Perfect score: 886  
Sequence: 1 MSSMPEYDYLFKLLIGDS.....EKSNNKIOSTPVKOSGGGCC 173

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCY\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	886	100.0	173	US-09-820-003a-2	Sequence 2, Appl1
2	866	97.1	222	US-09-820-003a-4	Sequence 4, Appl1
3	766.5	86.3	201	US-09-967-736-3	Sequence 3, Appl1
4	766.5	86.3	201	US-09-967-736-3	Sequence 8, Appl1
5	414	46.7	207	US-09-794-257-8	Sequence 5, Appl1
6	414	46.7	207	US-09-794-257-8	Sequence 8, Appl1
7	403.5	44.1	218	US-09-822-860-5	Sequence 1571, App
8	399.5	44.1	218	US-09-925-300-1571	Sequence 534, App
9	377.1	40.9	162	US-09-834-765-766	Sequence 16, App
10	362.5	40.9	198	US-09-945-173-5	Sequence 16, App
11	362.5	40.9	198	US-09-972-529-4	Sequence 4, Appl1
12	362.5	40.9	198	US-09-945-173-5	Sequence 4, Appl1
13	333	37.6	212	US-09-350-874-67	Sequence 67, Appl1
14	333	37.6	212	US-09-817-198A-4	Sequence 4, Appl1
15	321	36.2	212	US-09-817-198A-2	Sequence 2, Appl1
16	315.5	35.6	223	US-09-817-199A-4	Sequence 4, Appl1
17	314.5	35.5	223	US-09-817-199A-2	Sequence 2, Appl1
18	310	35.0	218	US-09-817-198A-5	Sequence 5, Appl1
19	292.5	33.0	191	US-09-794-257-14	Sequence 14, Appl1

20	292.5	33.0	191	US-10-051-986-3	Sequence 3, Appl1
21	292.5	33.0	312	US-09-925-302-783	Sequence 783, App
22	292	33.0	59	US-09-864-761-35038	Sequence 35038, A
23	267	30.1	216	US-09-945-173-10	Sequence 10, Appl1
24	258	29.1	239	US-09-925-301-1077	Sequence 1077, App
25	255	28.1	217	US-09-988-974-3	Sequence 9, Appl1
26	254	28.7	213	US-09-794-257-3	Sequence 8, Appl1
27	254	28.7	213	US-09-988-974-8	Sequence 45, Appl1
28	237.5	26.8	208	US-10-088-605-45	Sequence 45, Appl1
29	235.5	26.6	832	US-09-834-765-2	Sequence 1364, App
30	235	26.5	217	US-09-925-300-1364	Sequence 28, Appl1
31	219.5	24.8	183	US-09-765-288A-26	Sequence 2, Appl1
32	219.5	24.8	183	US-09-765-288A-26	Sequence 28, Appl1
33	219.5	24.8	189	US-10-104-484-4	Sequence 2, Appl1
34	216.5	24.4	189	US-10-104-484-2	Sequence 16, Appl1
35	214.5	24.2	189	US-10-067-813-16	Sequence 4, Appl1
36	212	23.9	190	US-09-850-373-4	Sequence 4, Appl1
37	212	23.9	624	US-09-834-765-762	Sequence 762, App
38	212	23.8	144	US-09-972-529-7	Sequence 7, Appl1
39	210.5	23.8	201	US-09-822-860-2	Sequence 2, Appl1
40	206	23.3	201	US-09-988-974-9	Sequence 5, Appl1
41	206	23.3	201	US-09-988-974-9	Sequence 9, Appl1
42	206	23.3	209	US-09-864-761-42996	Sequence 42996, A
43	205	23.1	168	US-09-834-765-765	Sequence 765, App
44	205	23.1	168	US-09-834-765-765	Sequence 629, App
45	200	22.6	208	US-09-925-302-629	

# ALIGNMENTS

```

RESULT 1
US-09-820-003a-2
; Sequence 2, Application US/09820003a
; Patent No. US20020142382A1
; GENERAL INFORMATION:
; APPLICANT: MERRUOY, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING, THESE HUMAN RAS-LIKE
; FILE REFERENCE: CU001196
; CURRENT FILING DATE: 2001-03-29
; CURRENT APPLICATION NUMBER: US/09/820,003A
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO. 2
; LENGTH: 173
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-820-003a-2

Query Match      100.0%   Score 886;   DB 10;   Length 173;
Best Local Similarity 100.0%;   Pred. No. 2.5e-80;
Matches 173;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY      1 MSSMPEYDYLFKLLIGDSGVKSCLLLRPADDTTSTSTIGVDFKIRITELDKTI 60
        |||
DB      1 MSSMPEYDYLFKLLIGDSGVKSCLLLRPADDTTSTSTIGVDFKIRITELDKTI 60
        |||

QY      61 KQIESFNNVKQWLOEIDRYASENVKLLVGNKCDLTKKYVDYTTAKFADSLGIPLE 120
        |||
DB      61 KQIESFNNVKQWLOEIDRYASENVKLLVGNKCDLTKKYVDYTTAKFADSLGIPLE 120
        |||

QY      121 TSAKNATNVEOSFMTMAELIKRMPGATAGAEKSNVKIOSTPVKOSGGGCC 173
        |||
DB      121 TSAKNATNVEOSFMTMAELIKRMPGATAGAEKSNVKIOSTPVKOSGGGCC 173
        |||

RESULT 2
US-09-820-003a-4
; Sequence 4, Application US/09820003a
; Patent No. US20020142382A1
; GENERAL INFORMATION:

```

APPLICANT: MERKULOV, Gennady et al.  
TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
FILE REFERENCE: C1001196  
CURRENT APPLICATION NUMBER: US/09/820, 003A  
CURRENT FILING DATE: 2001-03-29  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 222  
TYPE: PRT  
ORANISM: Homo sapien  
US-09-820-003a-4

Query Match 97.1%; Score 860; DB 10; Length 222;  
Best Local Similarity 84.4%; Pred. No. 1.2e-77;  
Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MSSMPEYDYLFLKLLIGDSGVKSCLLRFADDTYESYISTIGVDKIRITELDGKTI 60  
DB 18 MSSMPEYDYLFLKLLIGDSGVKSCLLRFADDTYESYISTIGVDKIRITELDGKTI 77  
QY 61 KLOI-----ESFNNVKOMLOEIDRYASENNKLL 88  
DB 78 KLOIMDTAGERRRTTSSYRGAGHIIYVDYTDQESFNNVKOMLOEIDRYASENNKLL 137  
QY 89 LVGNKCDLTTKKVVDYTTAKEFADSLGIPLETSAKNATNVEOSFMTMAEIKKRMGPAG 148  
DB 138 LVGNKCDLTTKKVVDYTTAKEFADSLGIPLETSAKNATNVEOSFMTMAEIKKRMGPAG 197  
QY 149 TAGGAEKSNVKIOSTPVKOSGGGCC 173  
DB 198 TAGGAEKSNVKIOSTPVKOSGGGCC 222

RESULT 3  
US-09-967-736-3  
Sequence 3, Application US/09967736  
Patent No. US20020103340A1  
GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.  
Lal, Preeti  
Corley, Neil C.  
Shah, Purvi  
TITLE OF INVENTION: RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/967, 736  
FILING DATE: 28-Sep-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/154,602  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0367 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: L1VPT0704  
CLONE: 2514506  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
US-09-967-736-3

Query Match 86.5%; Score 766.5; DB 10; Length 201;  
Best Local Similarity 77.2%; Pred. No. 1.8e-68;  
Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

QY 4 MNPEYDYLFLKLLIGDSGVKSCLLRFADDTYESYISTIGVDKIRITELDGKTIKLO 63  
DB 1 MNPEYDYLFLKLLIGDSGVKSCLLRFADDTYESYISTIGVDKIRITELDGKTIKLO 60  
QY 64 I-----ESFNNVKOMLOEIDRYASENNKLL 91  
DB 61 IMDTAGERRRTTSSYRGAGHIIYVDYTDQESFNNVKOMLOEIDRYASENNKLL 120  
QY 92 NKCDLTTKKVVDYTTAKEFADSLGIPLETSAKNATNVEOSFMTMAEIKKRMGPAGTAG 151  
DB 121 NKSDLTTKKVVDYTTAKEFADSLGIPLETSAKNATNVEOSFMTMAEIKKRMGPAGTAG 180  
QY 152 GAERKSNVKIOSTPVKOSGGGCC 173  
DB 181 GERKSNVKIOSTPVKOSGGGCC 201

RESULT 4  
US-09-967-736-8  
Sequence 8, Application US/09967736  
Patent No. US20020103340A1  
GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.  
Lal, Preeti  
Corley, Neil C.  
Shah, Purvi  
TITLE OF INVENTION: RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/967, 736  
FILING DATE: 28-Sep-2001  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/154,602  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0367 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 amino acids  
TYPE: amino acid  
STRANDEDNESS: single



TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 57006  
SEQUENCE DESCRIPTION: SEQ ID NO: 8  
US-09-967-736-8

Query Match 85.8%; Score 760.5; DB 10; Length 201;  
Best Local Similarity 76.2%; Pred. No. 6.9e-68;  
Matches 154; Conservative 7; Mismatches 8; Indels 33; Gaps 2;

QY 4 NMPEYDLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ 63  
DB 1 NMPEYDLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ 60  
QY 64 I-----ESFNNVKOMLOEIDRYASENVNKLIVG 91  
DB 61 IWDIAGGERFRTTAYYRGAMGIMLVYDITNKSFDNKKMIRNIEHSSADVERMLIG 120  
QY 92 NKCDLTTKRVVDTTAKFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGATAG 151  
DB 121 NKSDLTTRKVVDTTAKFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGAAAG 180  
QY 152 GAESNVKIQSTFVKSQSGGCC 173  
DB 181 G-ERNPLKIDSTFVKSASGCCC 201

RESULT 5  
US-09-794-257-8  
Sequence 8, Application US/09794257  
Patent No. US2002009804A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, No. US2002009804A1el  
FILE REFERENCE: 35800/209285  
CURRENT APPLICATION NUMBER: US/09/794,257  
CURRENT FILING DATE: 2001-02-27  
PRIOR APPLICATION NUMBER: 60/185,606  
PRIOR FILING DATE: 2000-02-29  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 207  
TYPE: PRT  
ORGANISM: homo sapiens  
US-09-794-257-8

Query Match 46.7%; Score 414; DB 10; Length 207;  
Best Local Similarity 44.2%; Pred. No. 9.7e-34;  
Matches 84; Conservative 31; Mismatches 37; Indels 38; Gaps 2;

QY 4 NMPEYDLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ 63  
DB 1 MAKTYDLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ 60  
QY 64 I-----ESFNNVKOMLOEIDRYASENVNKLIVG 91  
DB 61 IWDIAGGERFRTTAYYRGAMGIMLVYDITNKSFDNKKMIRNIEHSSADVERMLIG 120  
QY 92 NKCDLTTKRVVDTTAKFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGATAG 151  
DB 121 NKCDLTTKRVVDTTAKFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGATAG 180  
QY 146 PGATAGAEK 155  
DB 181 NSAGAGPVR 190

RESULT 6  
US-09-822-860-5  
Sequence 5, Application US/09822860

Patent No. US20020146795A1  
GENERAL INFORMATION:  
APPLICANT: ZHU, Shaoqing et al.  
TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
FILE REFERENCE: CL001214  
CURRENT APPLICATION NUMBER: US/09/822,860  
CURRENT FILING DATE: 2001-04-02  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 190  
TYPE: PRT  
ORGANISM: Discopyge ommata  
US-09-822-860-5

Query Match 46.6%; Score 413; DB 10; Length 190;  
Best Local Similarity 42.1%; Pred. No. 1.1e-33;  
Matches 80; Conservative 38; Mismatches 40; Indels 32; Gaps 1;

QY 9 DYLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ---- 64  
DB 1 DYLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ---- 60  
QY 65 -----ESFNNVKOMLOEIDRYASENVNKLIVG 96  
DB 61 GOERFRTTAYYRGAMGIMLVYDITNKSFDNKKMIRNIEHSSADVERMLIG 120  
QY 97 TTRKVVDTTAKFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGATAGAEKS 156  
DB 121 NERKQVSKRGEKLAIDYGIKPLETSKSNINVEAFITLARDIMTKIKMENSLOQA 180  
QY 157 NVKIQSTFVKS 166  
DB 181 VDLKSPVR 190

RESULT 7  
US-09-925-300-1571  
Sequence 1571, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1571  
LENGTH: 218  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-300-1571

Query Match 45.5%; Score 403.5; DB 10; Length 218;  
Best Local Similarity 39.7%; Pred. No. 1.1e-32;  
Matches 85; Conservative 32; Mismatches 30; Indels 67; Gaps 4;

QY 8 DYLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ---- 64  
DB 24 YDLFKLLIGDSGKSCLLRFADDDYTESYSTIGVDFKIRITELDGKTIKQ---- 83  
QY 65 -----ESFNNVKOMLOEIDRYASENVNKLIVG 95  
DB 84 AGGERFRTTAYYRGAMGIMLVYDITNKSFDNKKMIRNIEHSSADVERMLIG 143

QY	96	LTTKKVVDTYFAKAFASLIGIPETLETSAKNATINVEGSPMAAEIKRMGGATAGCAEK	155
Dd	144	MDDRVRVPRKGEGDIANEHGIRFETSANKININEKAFLTLAEDI	188
QY	156	SNVKISQPVKQ-----SGGS-----CC 173	
Dd	189	----LRKTVPKEPNSENVADISSGGGVGMKSCKC 218	

RESULT 8

```

US-09-925-302-534
Sequence 534, Application US/09925302
Patent No. US20020044941A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA104
CURRENT APPLICATION NUMBER: US/09/925,302
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05918
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 896
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 534
LENGTH: 246
TYPE: prt
ORGANISM: Homo sapiens
US-09-925-302-534

```

Query Match	44.1%;	Score 390.5;	DB 10;	Length 246;
Best Local Similarity	41.7%;	Pred. No. 2.5e-31;		
Matches 78; Conservative	35;	Mismatches 41;	Indels 33;	Gaps 2

[illegible]

## RESULT 9

US-09-834-765-766  
Sequence 766, Application US/09834765  
Patent No. US20020055478A1  
GENERAL INFORMATION:  
APPLICANT: Mary Faris  
APPLICANT: Pia M. Challita-Eid  
APPLICANT: Arthur B. Raitano  
APPLICANT: Steve Chappell Mitchell  
APPLICANT: Daniel E. H. Afar  
APPLICANT: Aya Jakobovits  
TITLE OF INVENTION: GTP-BINDING PROTEIN USEFUL IN TREATMENT  
TITLE OF INVENTION: AND DETECTION OF CANCER  
FILE REFERENCE: 129.6US01  
CURRENT APPLICATION NUMBER: US/09/834,765  
CURRENT FILING DATE: 2001-09-21  
PRIOR APPLICATION NUMBER: 60/197,647  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 770  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 766

```

; LENGTH: 162
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-834-765-766

```

Query Match	41.98; Score 371; DB 10; Length 162;
-------------	--------------------------------------

Matches 73; Conservative 28; Mismatches 29; Indels 32; Gaps 1;

## QY 12

[illegible]

RESULT 10

```

US-09/794-257-16
; Sequence 16, Application US/09794257
; Patent No. US20020009804A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, NO. US20020009804A1e1
; TITLE OF INVENTION: Human G-Proteins
; FILE REFERENCE: 35800/209285
; CURRENT APPLICATION NUMBER: US/09/794,257
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/185,606
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 198
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Pfam accession number PF00071
US-09-794-257-16

```

Query Match	40.98;	Score 362.5;	DB 10;	Length 198;
Best Local Similarity	41.7%;	Pred. No. 1.1e-28;		
Matches	86;	Conservative	25;	Mismatches 40;
				Indels 55;
				Gaps 6;

```

Qy 13 KLIIIGSGVKSCLLRREADDTTESTISITGVDFPIRTLEIDGKIRIKOI----- 64
Db 1 KLVLIGDSGVKSSLLIRFDNKFVEEYIPIITGVDFYTKVEYDGKTVKLIQIMDTAQGR 60
Qy 65 -----ESEFNWYKQWLOEIDRYA--SENNKRLVGNKCOLT- 97
Db 61 FRALRPAYRGAGQFLVLYIDITSRDSFNWKKWMEILLRADKXENPIYLVGNKCOLLED 120
Qy 98 -----TKKVVDDYTTAKEFADSLG-IPLETSAAKNATNVEOSEMYTAAEIIKRRMGP 147
Db 121 DEDLEITFEGQRRVYSTEGEGLAAKEIHALPMEITSAKTNNVVEEAFELAREILIKV--- 177
Qy 148 ATAGAAGKSNYKIDSTPYKQSGGGCC 173
Db 178 -----SEVNNVNDLO-PAKKKKSKCC 196

```

## RESULT 11

US-09-945-173-5  
; Sequence 5, Application US/09945173  
; Patent No. US20020127568A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc  
; APPLICANT: Meyers, Rachel

```

: TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
: FILE REFERENCE: 38155-20035.00
: CURRENT APPLICATION NUMBER: US/09/945,173
: PRIOR FILING DATE: 2001-08-31
: PRIOR APPLICATION NUMBER: US 60/229,293
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 198
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus amino acid sequence
US-09-945-173-5

Query Match
Best Local Similarity 40.9%; Score 362.5; DB 10; Length 198;
Matches 86; Conservative 25; Mismatches 40; Indels 55; Gaps 6;

QY 13 KLLLGSGVSKCLLRFPADDTYESYISTGVDFKRTIELDKTKIKQI----- 64
DB 1 KVLIGDSGVSKSLIRPTDKRVEEYIPTLVDFYTRIVEDGKIVKQIWDTAGOER 60
QY 65 -----ESTNNYKQWLOEIDRYA--SENVNKLIVGNKCDLT- 97
DB 61 FALRPAYYRGAGFLVYDITSRDSFENKMKLEILRHADKDNVPIVLGNKCDLED 120
QY 98 -----TKKVYDYTTAKEPADSLG-IPLETSAKNATNVEQSFMTMAEIKRRMGPG 147
DB 121 DEDLELTGOKRVVSTEEGALAKELGALPDMETSAKNTNVEAFELAREILKVV--- 177
QY 148 ATAGAESKNVYKIQSTPVKQSGGCC 173
DB 178 -----SEVNVNLDQ-PAKKKSKCC 196

RESULT 12
US-09-972-529-4
: Sequence 4, Application US/09972529
: Patent No. US20020150916A1
: GENERAL INFORMATION:
: APPLICANT: Millennium Pharmaceuticals, Inc.
: TITLE OF INVENTION: 47316, A NOVEL HUMAN G-PROTEIN AND USES
: FILE REFERENCE: 38155-20041.00
: CURRENT APPLICATION NUMBER: US/09/972,529
: PRIOR FILING DATE: 2002-04-08
: PRIOR APPLICATION NUMBER: US 60/237,716
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 198
: TYPE: PRT
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Consensus amino acid sequence
US-09-972-529-4

Query Match
Best Local Similarity 40.9%; Score 362.5; DB 10; Length 198;
Matches 86; Conservative 25; Mismatches 40; Indels 55; Gaps 6;

QY 13 KLLLGSGVSKCLLRFPADDTYESYISTGVDFKRTIELDKTKIKQI----- 64
DB 1 KVLIGDSGVSKSLIRPTDKRVEEYIPTLVDFYTRIVEDGKIVKQIWDTAGOER 60
QY 65 -----ESTNNYKQWLOEIDRYA--SENVNKLIVGNKCDLT- 97
DB 61 FALRPAYYRGAGFLVYDITSRDSFENKMKLEILRHADKDNVPIVLGNKCDLED 120
QY 98 -----TKKVYDYTTAKEPADSLG-IPLETSAKNATNVEQSFMTMAEIKRRMGPG 147
DB 121 DEDLELTGOKRVVSTEEGALAKELGALPDMETSAKNTNVEAFELAREILKVV--- 177
QY 148 ATAGAESKNVYKIQSTPVKQSGGCC 173
DB 178 -----SEVNVNLDQ-PAKKKSKCC 196

```

```

QY 98 -----TKKVYDYTTAKEPADSLG-IPLETSAKNATNVEQSFMTMAEIKRRMGPG 147
DB 121 DEDLELTGOKRVVSTEEGALAKELGALPDMETSAKNTNVEAFELAREILKVV--- 177
QY 148 ATAGAESKNVYKIQSTPVKQSGGCC 173
DB 178 -----SEVNVNLDQ-PAKKKSKCC 196

RESULT 13
US-09-350-874-67
: Sequence 67, Application US/09350874
: Patent No. US20020019020A1
: GENERAL INFORMATION:
: APPLICANT: Rhodes, Kenneth
: TITLE OF INVENTION: METHODS FOR TREATING CARDIOVASCULAR DISORDERS
: FILE REFERENCE: NMT-069
: CURRENT APPLICATION NUMBER: US/09/350,874
: PRIOR FILING DATE: 1999-07-09
: EARLIER APPLICATION NUMBER: USSN 60/110,277
: EARLIER FILING DATE: 1998-11-30
: EARLIER APPLICATION NUMBER: USSN 60/110,033
: EARLIER FILING DATE: 1998-11-25
: EARLIER APPLICATION NUMBER: USSN 60/109,333
: EARLIER FILING DATE: 1998-11-20
: EARLIER APPLICATION NUMBER: USSN 09/298,731
: NUMBER OF SEQ ID NOS: 72
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 67
: LENGTH: 212
: TYPE: PRT
: ORGANISM: Rattus sp.
US-09-350-874-67

Query Match
Best Local Similarity 37.6%; Score 333; DB 10; Length 212;
Matches 73; Conservative 38; Mismatches 55; Indels 44; Gaps 2;

QY 8 YDLFLLIGSGVSKCLLRFPADDTYESYISTGVDFKRTIELDKTKIKQI--- 64
DB 3 YALFYIITDPTGVKSCILLQFTDKRFQPVHDLTIGVEFGARMITIDKQIKQIWDY 62
QY 65 -----ESTNNYKQWLOEIDRYA--SENVNKLIVGNKCDLT- 97
DB 63 AGESRSTSTSYRGAAGLALVYDITRDTFNHLLTLEDAKSHSNMAYIMLGNKSD 122
QY 96 LTRKVVYDYTTAKEPADSLGIPLETSAKNATNVEQSFMTMAEIKRRMGPGATAGAEK 155
DB 123 LESRREVKKREGEAFAREHGLIMETSAKNTNVEAFINTAKETKIEQGVFDINNEA 182
QY 156 SNVKIQSTPV-----KQSGGCC 173
DB 183 NGIKIGPOHAATNASHGNGGQOAGGCC 212

RESULT 14
US-09-817-198A-4
: Sequence 4, Application US/09817198A
: Patent No. US20020146758A1
: GENERAL INFORMATION:
: APPLICANT: YE, Jane et al.
: TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
: FILE REFERENCE: CL001188
: CURRENT APPLICATION NUMBER: US/09/817,198A
: PRIOR FILING DATE: 2001-03-27
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 4

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 17, 2002, 13:05:13 ; Search time 2644 Seconds  
(without alignments)  
1904.230 Million cell updates/sec

Title: US-09-820-003A-2

Perfect score: 886  
Sequence: 1 MSSMPEYDYLFKLLIGDS.....EKSNNKIOSTPYKOSGGGCC 173

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n model -DEV=xlh  
-O=/cgn2.1/USPFO.spool/US09820003/runat.13112002.134616.3176/app.query.fasta.1.327  
-DB=Genembl -OPMT=fastap -SUPER=rgc -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40 cdi -LIST=45  
-DOCFALT=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USFR=US09820003.ecgn1.1.1616.ernat.13112002.134616.3176 -NCPU=6 -ICPU=3  
-NO.XLPXY -NO.MMAP -LARGEDUERY -NEG SCORES=0 -MATT -LONCLOG -DEV.TIMEOUT=120  
-WARN.TIMEOUT=10 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Genembl:\*  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pa:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pla:\*  
35: em\_hlg\_rod:\*  
36: em\_hlg\_mam:\*  
37: em\_hlg\_vtl:\*  
38: em\_sy:\*  
39: em\_hlgo\_hum:\*  
40: em\_hlgo\_mus:\*  
41: em\_hlgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	860	97.1	615	6	E03861
2	860	97.1	618	9	AF498929
3	860	97.1	723	9	H0WRAB1A
4	860	97.1	840	4	CFRAB1
5	860	97.1	1428	10	MMYPT1
6	860	97.1	1433	9	BC000905
7	860	97.1	1444	10	BC002077
8	860	97.1	2686	10	AF226873
9	849	95.8	815	10	RATRASA
10	838	94.6	841	5	DYGORAB1
11	768.5	86.7	925	6	AR070365
12	768.5	86.7	1985	6	AX086155
13	768.5	86.7	1985	9	HSMB01608
14	762.5	86.1	654	10	RMRAB1B
15	762.5	86.1	161955	2	AC106368
16	761.5	85.9	1859	10	BC016408
17	758.5	85.6	151988	2	AC123143
18	740.5	83.6	1988	2	HSB245875
19	737.5	82.8	185285	2	AC094411
20	733.5	82.8	169095	2	AL133944
21	733.5	82.8	169095	2	AL133944
22	733.5	82.8	207945	2	AC117841
23	733.5	82.7	1778	3	LSRAB1
24	732	81.3	1066	10	RH063023
25	732	81.3	174028	2	AC093834
26	730	79.0	2257	9	AK035927
27	683	77.1	1772	3	DB4312
28	671	75.7	952	3	AF013572
29	658.5	74.3	1868	3	CNS0198R
30	655	73.9	1868	3	AY121695
31	645	72.8	979	8	LJRB1C
32	644	72.7	938	8	AY052204
33	644	72.7	938	8	AY052204
34	636.5	71.8	864	8	PEAGTBP10
35	634.5	71.6	609	8	AY060495
36	633	71.4	636	8	CNS01BW9
37	633	71.4	640	8	AY114013
38	633	71.4	931	8	AF324990
39	633	71.4	932	8	AY085384
40	633	71.4	939	8	AY080753
41	632.5	71.4	1008	8	NP43GTP
42	632.5	71.3	1074	8	AB024994
43	630.5	71.2	1129	8	LJRB1D
44	628.5	70.9	898	8	NTRGBP
45	626.5	70.7	860	8	LJRB1E

RESULT 1

#### ALIGNMENTS

E03861 LOCUS E03861 615 bp RNA linear PAT 29-SEP-1997  
 DEFINITION cDNA encoding GTP binding protein ypt1.  
 ACCESSION E03861  
 VERSION E03861.1 GI:2172075  
 KEYWORDS UP 1992226000-A/3.  
 SOURCE Mus sp.  
 ORGANISM Mus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.  
 REFERENCE 1 (bases 1 to 615)  
 AUTHORS Yamaguchi, K., Uchida, K., Takai, Y., Aisaka, K. and Sakurada, K.  
 TITLE LOW-MOLECULAR WEIGHT PROTEIN BINDING TO GUANOSINE TRIPHOSPHATE  
 JOURNAL Patent: JP 1992226000-A 3 14-AUG-1992;  
 KYOMA HAKKO KOGYO CO LTD  
 COMMENT OS Mus sp (mouse)  
 PN JP 1992226000-A/3  
 PD 14-AUG-1992  
 PF 25-APR-1991 JP 1991095304  
 PR 21-MAY-1990 JP 90P 130547  
 PI YAMAGUCHI KAZUO, UCHIDA KAZUHISA, TAKAI YOSHIMI, AISAKA KAZUO,  
 PI SAKURADA KAZUHIRO  
 PC C07K15/18,C07K13/00,C07K15/06,C12N15/12,G01N33/50//A61K37/02;  
 CC strandeness: Single;  
 topology: Linear;  
 location/Qualifiers  
 1..615  
 /organism="Mus sp."  
 /db\_xref="taxon:10095"  
 BASE COUNT 196 a 117 c 153 g 149 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.53e-75 Length: 615  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 Gaps: 1

US-09-820-003a-2 (1-173) x E03861 (1-615)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 |||||  
 DB 1 ATGTCAGCATGAAATCCCGAATATGATTAATTCAGATTACTTGATGGCGATTCT 60  
 |||||  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 |||||  
 DB 61 GGGGTGGAAAGTCCCTGCTCTCTTAGTTGCGAGTGAATACGATATACGGAAGCTAC 120  
 |||||  
 QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
 |||||  
 DB 121 ATCAGCAAAATGGTGTGATTTCAAGATACGAACTATAGATTAGTGGGAAAAAATC 180  
 |||||  
 QY 61 LysLeuGlnIle----- 64  
 |||||  
 DB 181 AAGCTACAGATATGGACACAGACAGCCAGGAAGATTTCGAACAATCACTTCAGATTAT 240  
 |||||  
 QY 65 -----GluSerPheAsn 68  
 |||||  
 DB 241 TACAGAGAGCCCATGGCATCATAGTGTGTATGATGTGACAGATACAGAGATCCCTTCAT 300  
 |||||  
 QY 69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
 |||||  
 DB 301 AACGTTAAACAGTGGCTCCAGAGATAGATCGCTACGCCAATGAAATATGTAACAAGTTG 360  
 |||||  
 QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
 |||||  
 DB 361 TTGGTAGGAGCAAAATGTGACCTGCACCAAAAGAAAGTAGTAGACTACACAACAGCAAG 420  
 |||||  
 QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluTyrSerAlaLysAsnAlaThrAsn 128  
 |||||  
 DB 421 GAATTTGACAGATCCCTTGGAATTCATTTTGGAAACCACTGCTAAGAACGCAACGAAAT 480  
 |||||

QY 129 ValGluInSerPheMetThrMetAlaGluIleLysLysArgMetGlyProGlyAla 148  
 |||||  
 DB 481 GTAGACAGCTTTTCATCATGACATGCGACGCTGAGTTAAAGCAATGGCTCCGAGAGCT 540  
 |||||  
 QY 149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 |||||  
 DB 541 ACACCTGCTGCTGCCGAGAAAGTCCCAATGTTAAATCCAGAGCACTCCAGTCACAGCTCA 600  
 |||||  
 QY 169 GlyGlyGlyCysCys 173  
 |||||  
 DB 601 GGTGAGGCTCTGCTCC 615  
 |||||

RESULT 2  
 AF498929  
 LOCUS AF498929  
 DEFINITION 618 bp mRNA linear PRI 01-MAY-2002  
 cds.  
 ACCESSION AF498929  
 VERSION AF498929.1 GI:20379033  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 618)  
 Puhl, H.L. Iii, Ikeda, S.R. and Aronstam, R.S.  
 TITLE Homo sapiens RAB family small GTP binding protein RAB1A  
 JOURNAL Unpublished  
 2 (bases 1 to 618)  
 Puhl, H.L. Iii, Ikeda, S.R. and Aronstam, R.S.  
 REFERENCE Direct Submission  
 TITLE Submitted (05-APR-2002) CDNA Resource Center, Guthrie Research  
 AUTHORS Institute, One Guthrie Square, Sayre, PA 18840, USA  
 JOURNAL location/Qualifiers  
 1..618  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="brain"  
 1..618  
 /gene="RAB1A"  
 1..618  
 /gene="RAB1A"  
 /note="RAB family member"  
 /codon\_start=1  
 /product="small GTP binding protein RAB1A"  
 /protein\_id="P44710.1"  
 /db\_xref="GI:20379034"  
 /translation="MSSMNPEDYVLEKLLIGSGVGSCLLRFADDTYESISTI  
 GVDPKRTIILDEKTIKLIQIMDPAQGERFRTISYYRGAGIIVYDVYDQESFNVY  
 KQMLQEDRYASENVNKLIVGNCQDLTKKVDYVYTAKEPADSLGIPFLETSKRNATN  
 VDSFPMMAEIKRMGPGATAGAEISNVKISTPKVGGGCC"

BASE COUNT 203 a 110 c 149 g 156 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.53e-75 Length: 618  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 Gaps: 1

US-09-820-003a-2 (1-173) x AF498929 (1-618)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 |||||  
 DB 1 ATGTCAGCATGAAATCCCGAATATGATTAATTCAGATTACTTGATGGCGACTCA 60  
 |||||  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 |||||  
 DB 61 GGGGTGGAAAGTCCCTGCTCTCTTAGTTGCGAGTGAATACGATATACGGAAGCTAC 120  
 |||||

```

OY 41 IleserThrIleGlyValAlaSpPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60
Db 121 ATCAGCAACATTCGTGTGATTCATAATAGACTATAGAGTTAGACGGAAACATTC 180
OY 61 LysLeuGlnIle----- 64
Db 181 AACGCTCAATATGGACACAGACGCCAGGAAAGATTCCACATCCACCTCCAGTTAT 240
OY 65 -----GluSerPheAsn 68
Db 241 TACAGAGAGCCCATGCAATCATGTTGTTATGATGCACATCAGAGAGTCCCTTCAT 300
OY 69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88
Db 301 AATGTTAAACAGGCTGCGAGAAATAGATCGTTATGCAAGTGAATGTCACAAATG 360
OY 89 LeuValGlyAsnLysCysAspLeuThrThrLysValValAspTyrThrAlaLys 108
Db 361 TTGGTAGGGAACAATGATGATCCATCCACCAAGAAAGTAGTAGACACAAACAGCAG 420
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluTrpSerAlaLysAsnAlaThrAsn 128
Db 421 GAATTTGCTGATTCCTTGGAATTCGCTTTTGGAACACAGTGTCAAGATCAACGAA 480
OY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148
Db 481 GTAGAACAGCTTTCATGACATGCGAGTGAATTAAGCAATGGTCCCGAGACA 540
OY 149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168
Db 541 ACAGCTGGTGTGCTGAGACATGCAATGTTAAATTCAGACACTCCAGTCAAGCAGTCA 600
OY 169 GlyGlyGlyCysCys 173
Db 601 GGTGGAGGTGCTGC 615

RESULT 3
HUMRABIA 723 bp mRNA linear PRI 08-JAN-1995
LOCUS Homo sapiens GTP-binding protein (RAB1) mRNA, complete cds.
DEFINITION M28209 J04941
VERSION M28209.1 GI:550059
KEYWORDS GTP-binding protein; ras oncogene.
SOURCE Homo sapiens (tissue library: of J.Mallet) pheochromocytoma CDNA to
        mRNA.
ORGANISM Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 723)
AUTHORS Zahraoui, A., Touchot, N., Chardin, P. and Tavilian, A.
TITLE The human Rab genes encode a family of GTP-binding proteins related
        to yeast Ypt1 and Sec4 products involved in secretion
JOURNAL J Biol Chem. 264 (21), 12394-12401 (1989)
MEDLINE 8930866
PUBMED 2501306
FEATURES
        Source
        on Sep 27, 1994 this sequence version replaced gi:341862.
        Location/Qualifiers
            1..723
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="Unassigned"
            /tissue_type="pheochromocytoma"
            /tissue_lib="of J.Mallet"
            1..723
            /gene="RAB1"
            51..668
            /gene="RAB1"
            /codon_start=1
            /product="GTP-binding protein"
            /protein_id="AA60240.1"
            /db_xref="GI:550060"
            /db_xref="GDB:G00-118-857"
            /translation="MSSMNPEDYLFKLLIGDSGVGRSCLLAFADDTESYLSIT1

CUDFKRTIEIDCKTITQIDNTAQGERETTSYRGACITIVYDVTDOSENNV
KQNGIDIRYSERVNKLIVGNKCDLITKRVYDITTAKEFASGJIFLETISKNTIN
VEQSEFIMAEIKRMGPATRGAGKSNVYDITSTIVKOSGGCC"

BASE COUNT      225 a      142 c      180 g      176 t
ORIGIN
Alignment Scores:
Pred. No.:      5,46e-75
Score:          860.00
Percent Similarity: 84.39%
Best Local Similarity: 84.39%
Query Match:    97.07%
DB:              9
Gaps:            1

US-09-820-003a-2 (1-173) x HUMRABIA (1-723)
OY 1 MetSerSerMetAsnProGluTrpAspTyrIleuPheLysLeuLeuIleGlyAspSer 20
Db 51 ATGTCAGCATGAATCCCGAATATATTATTATTCAGTTCATCTTGATGGCCGACTCA 110
OY 21 GlyValGlyLysSerCysLeuLeuAspPheAlaAspAspThrTyrThrGluSerTyr 40
Db 111 GGGCTTGGAAGACTTCTCCTCTCTTCTTGTAGTTGCGAGATGATATATACGAAAGCTAC 170
OY 41 IleserThrIleGlyValAlaSpPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60
Db 171 ATCAGCACAAATGTGTGATTCATAAATAAGAACTATAGACTAGACGGGAAACAAATC 230
OY 61 LysLeuGlnIle----- 64
Db 231 AACGCTCAATATATGGACACAGCAGCAGGAAAGATTTCGACATCACCTCCAGTTAT 290
OY 65 -----GluSerPheAsn 68
Db 291 TACAGAGAGCCCATGCAATCATAGTTGTATGATGTGACATCAGAGAGTCTTCAT 350
OY 69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88
Db 351 AATGTTAAACAGCTGCGAGAAATAGATGCTTATGCCAGTGAATGTCACAAATG 410
OY 89 LeuValGlyAsnLysCysAspLeuThrThrLysValValAspTyrThrAlaLys 108
Db 411 TTGGTAGGGAACAATGATGATCCATCCACCAAGAAAGTAGTAGACTCACACAGCAG 470
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluTrpSerAlaLysAsnAlaThrAsn 128
Db 471 GAATTTGCTGATTCCTTGGAATTCGCTTTTGGAACACAGTGTCAAGATCGAACGAC 530
OY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148
Db 531 GTAGAACAGCTTTCATGACATGCGAGTGAATTAAGCAATGGTCCCGAGACA 590
OY 149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168
Db 591 ACAGCTGGTGTGCTGAGACATGCAATGTTAAATTCAGACACTCCAGTCAAGCAGTCA 650
OY 169 GlyGlyGlyCysCys 173
Db 651 GGTGGAGGTGCTGC 665

RESULT 4
CERAB1 840 bp mRNA linear MAM 30-SEP-1999
LOCUS Canine rab1 mRNA for ras-related GTP-binding protein involved in
        membrane traffic.
DEFINITION X56384
VERSION X56384.1 GI:913
KEYWORDS GTP-binding protein; protein transport; ras gene; ras gene family.
SOURCE Canis familiaris.
ORGANISM Canis familiaris
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 840)

```

AUTHORS Chavrier, P., Vingron, M., Sander, C., Simons, K. and Zerial, M.  
 TITLE Molecular cloning of YPT1/SEC4-related cDNAs from an epithelial cell line  
 JOURNAL Mol. Cell. Biol. 10 (12), 6578-6585 (1990)  
 MEDLINE 91061765  
 PUBMED 2123294  
 REFERENCE 2 (bases 1 to 840)  
 AUTHORS Zerial, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-AUG-1990) Zerial M., EMBL, Meyerhofstrasse 1, 6900 Heidelberg, FRG

FEATURES  
 source  
 Location/Qualifiers  
 1..840  
 /organism="Canis familiaris"  
 /strain="cocker spaniel"  
 /db\_xref="taxon:9615"  
 /clone="47"  
 /cell\_line="MDCK"  
 149..766  
 /gene="rab1"  
 149..766  
 /gene="rab1"  
 /function="GTP binding protein"  
 /codon\_start=1  
 /protein\_id="CAB56775.1"  
 /db\_xref="GI:6006435"  
 /translation="MSSMNPEDYDLFKLLIGDSGVKSCLLRFADDTYESISTI  
 GVDFKRTIEDLCKTKIKLOIWDPAOGREFRTISSYRGHGIIVYDWDQSFNNV  
 KQWLOEIDRASENVNKLIVGNKCDLTKKVVYDTTAKKEADSLGIPLETSSKNATN  
 VEOSFMTMAEIKRMKPGATAGAAENSVNKIOSTPVKOSGGGCC"

BASE COUNT 250 a 176 c 220 g 194 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 6.52e-75 Length: 840  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 DB: 4 Gaps: 1

US-09-820-003a-2 (1-173) x CFRAB1 (1-840)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 |||||  
 Db 149 ATGTCAGCAGATCCCGATATGATTATTATCAAGTACTTCGATGGCGACT 208  
 |||||  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 |||||  
 Db 209 GGGCTTGGAAGTCTTCCTCTCTAGGTTTCAGATGATACATATACAGAAAGCTAC 268  
 |||||  
 QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGlyLeuAlaAspGlyLysThrIle 60  
 |||||  
 Db 269 ATCGGCCCAATGGTGGTGGATTTCAAAATAGAACTATAGATTAGATGGGAAAAACAATC 328  
 |||||  
 QY 61 LysLeuGlnIle----- 64  
 |||||  
 Db 329 AAGCTTCAAAATATGGACACACAGAGGCCCAAGAAAGATTTCGAACATCACTCCAGTTAT 388  
 |||||  
 QY 65 -----GluSerPheAsn 68  
 |||||  
 Db 389 TACAGAGAGCCCATGCATCATAGTGTGTATGACGTACAGATCAGAGAGTCTTCAT 448  
 |||||  
 QY 69 AsnValLysGlnTyrPheGlnGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
 |||||  
 Db 449 AATCTTAACAGTGGTGGCAGGAATAGACCCTTATGCGACGTGAAAACGTCACACAGCTTG 508  
 |||||  
 QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
 |||||  
 Db 509 TTGGTAGGAAACAATGCGATCTGACCAACAAGAAAGTAGAGCTACACACAGCAAG 568  
 |||||  
 QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
 |||||

Db 569 GAATTTGGCGGATTCCTTGGAATTCATTTTGGAAACCACTGCTTAAGACCAACAAT 628  
 QY 129 ValGlnSerPheMetThrMetAlaGlnIleLysLysArgMetGlyProGlyAla 148  
 |||||  
 Db 629 GTAGACAGCTTTCATGACGATGCGACGTGAGTTAAAAAGCAATGGGCTCGAGCA 688  
 |||||  
 QY 149 ThrAlaGlyAlaGlnLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 |||||  
 Db 689 ACACGTGGTGGTCAGAGAACTCAATGTGAAATTCAGAGCACTCCGCTCAACAGATCA 748  
 |||||  
 QY 169 GlyGlyGlyCysCys 173  
 |||||  
 Db 749 GGTGAGGTGCTGC 763

RESULT 5  
 MMYPT1 1428 bp mRNA linear ROD 12-SEP-1993  
 LOCUS Mouse mRNA for ras-related YPT1 protein.  
 ACCESSION Y00094  
 VERSION Y00094.1 GI:55456  
 KEYWORDS ras-related protein; YPT1 gene.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1428)  
 Gallowitz, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-NOV-1987) Gallowitz D., Max-Planck-Institut fuer Biophysikalische Chemie, Abt. Molekulare Genetik, Postfach 2841, D-3400 Goettingen, FRG  
 2 (bases 1 to 1428)  
 Haubruck, H., Disela, C., Wagner, P. and Gallowitz, D.  
 REFERENCE The ras-related ypt1 protein is an ubiquitous eukaryotic protein: isolation and sequence analysis of mouse cDNA clones highly homologous to the yeast YPT1 gene  
 EMBO J. 6 (1987) in press  
 Location/Qualifiers  
 1..1428  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_line="F9 and C3H10T1/2"  
 130..747  
 /note="Ypt1 protein (AA 1-205)"  
 /codon\_start=1  
 /protein\_id="CAA68284.1"  
 /db\_xref="GI:55457"  
 /db\_xref="MSD:WGI:97842"  
 /db\_xref="SWISS-PROT:P11476"  
 /translation="MSSMNPEDYDLFKLLIGDSGVKSCLLRFADDTYESISTI  
 GVDFKRTIEDLCKTKIKLOIWDPAOGREFRTISSYRGHGIIVYDWDQSFNNV  
 KQWLOEIDRASENVNKLIVGNKCDLTKKVVYDTTAKKEADSLGIPLETSSKNATN  
 VEOSFMTMAEIKRMKPGATAGAAENSVNKIOSTPVKOSGGGCC"  
 1401..1406  
 /note="pol. polyA signal"  
 1428  
 /note="pol. polyA site"  
 polyA\_site  
 BASE COUNT 410 a 263 c 324 g 431 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.22e-74 Length: 1428  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 DB: 10 Gaps: 1

US-09-820-003a-2 (1-173) x MMYPT1 (1-1428)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 |||||  
 Db 130 ATGTCAGCAGATCCCGATATGATTATTATCAAGTACTTCGATGGCGACTTCT 189





|||||  
Db 659 GTAAACAGCTTTTCACGACGATGCGACCTAGATTAAAGCCAAATGGTCCCGGAGCA 718  
Oy 149 ThrAAGlGyAlAGlulYsSerAsnValYsIleGlnSerThrProValYsGlnSer 168  
Db 719 ACACCTGCTGCTGCGAGAAAGTCCAAATGTTAAATTCAGACACTCCAGTCAAGCAGTCA 778  
Oy 169 GlyIleGlyCysCys 173  
Db 779 GGTGAGGCTCTGCTGC 793  
RESULT 7  
BC002077  
LOCUS BC002077 1444 bp mRNA linear ROD 07-AUG-2002  
DEFINITION Mus musculus, RAB1, member RAS oncogene family, clone MGC:6226  
IMAGE:3592802, mRNA, complete cds.  
ACCESSION BC002077  
KEYWORDS BC002077.1 GI:12805232  
SOURCE MGC.  
ORGANISM house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1444)  
AUTHORS Strausberg, R.  
JOURNAL Direct Submission  
Submitted (31-JAN-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
REMARK Contact: MGC help desk  
COMMENT Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdepaxil.stanford.edu](mailto:mcdepaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 11 Row: a Column: 2  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 6679586.  
FEATURES  
Source  
1.1444  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/map="FVB/N"  
/clone="MGC:6226 IMAGE:3592802"  
/tissue\_type="Mammary tumor. Metastatic adenoma. TGF alpha  
model. 10 month old virgin mouse. Taken by biopsy."  
/clone\_lib="NCI\_CGAP\_Mam1"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6"  
196..813  
CDS  
/codon\_start=1  
/product="RAB1, member RAS oncogene family"  
/protein\_id="AAH02077.1"  
/db\_xref="GI:12805233"  
/db\_xref="locusid:5861"  
/translation="MSSMNPEDYLEKLLIGSGVKSCLLRPADDTYESYIST  
GVDPKTRITLDGKTKLQIWDIAGQERFTTTSYRGAGIIVYDVTDQSEFNW  
KWLQIEDIRASENVKLLVGNKCDLTTKKVVDYTTAKERADSLGIPELTSAKNATN  
VEOSFTMAEIKRMGPGATGAGAKSNVKIOSTPVKOSGGGCC"  
BASE COUNT 412 a 270 c 342 g 420 t  
ORIGIN  
Alignment Scores:  
47

Pred. No.: 1.24e-74 Length: 1444  
Score: 860.00 Matches: 173  
Percent Similarity: 84.39% Conservative: 0  
Best local Similarity: 84.39% Mismatches: 0  
Query Match: 97.07% Indels: 32  
DB: 10 Gaps: 1  
US-09-820-003a-2 (1-173) x BC002077 (1-1444)  
Oy 1 MetSerSerMetAnpProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
Db 196 ATGTCCAGCATGAATCCCGAATATGATTATTATTCAGTACTTCTGATTGGCGATTCT 255  
Oy 21 GlyValGlyIysSerCysLeuLeuLeuArPheAlaAspAspThrTyrThrGluSerTyr 40  
Db 256 GGGTGTGGAAGTCTCTGCTCTCTCTTACGTTTGACATGATFACGGAAGCTAC 315  
Oy 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
Db 316 ATCAGCACAAATGGTGTGATTCAGATACGAACCTATAGATTAGATGGGAAAAACATC 375  
Oy 61 LysLeuGlnIle-----  
Db 376 AAGCTACAGATATATGGACACAGACGACGCGAAGATTTCGAACATTCATTCAATTAT 435  
Oy 65 -----GluSerPheAsn 68  
Db 436 TACAGAGAGCCCATGGCATCATGTTGTGTATGATGACACAGATCAGAGATCCTTCAT 495  
Oy 69 AsnValIysGlnTTPLeuGlnIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
Db 496 AACGTTAAACAGTGGCGCAGAGATGATCGCTACGCCAGTGAATAATGTCACAAAGTTG 555  
Oy 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValAlaAspTyrThrThrAlaLys 108  
Db 556 TTGGTAGGGAACAATGTGACCTACCAACAAGAAAGTAGAGCTACACAGACGCAAG 615  
Oy 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
Db 616 GAATTCGAGATTCCTCTGGAAATTCATTTTGGAAACCAAGTGTAAAGCAACGCAACAT 675  
Oy 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysAspMetGlyProGlyAla 148  
Db 676 GTAAACAGCTTTTCACGACGATGCGACCTGAGATTAAAGCCAAATGGGCTCGGAGCT 735  
Oy 149 ThrAAGlGyAlAGlulYsSerAsnValYsIleGlnSerThrProValYsGlnSer 168  
Db 736 ACACCTGCTGCTGCGAGAAAGTCCAAATGTTAAATTCAGACACTCCAGTCAAGCAGTCA 795  
Oy 169 GlyIleGlyCysCys 173  
Db 796 GGTGAGGCTCTGCTGC 810  
RESULT 8  
AF226873  
LOCUS AF226873 2686 bp mRNA linear ROD 13-FEB-2000  
DEFINITION Mus musculus small GTP-binding protein RAB1A mRNA, complete cds.  
ACCESSION AF226873  
VERSION AF226873.1 GI:6969621  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2686)  
AUTHORS Wu, G. and Dorn, G.W. II.  
JOURNAL Mouse RAB1A, member of RAS oncogene family mRNA  
Unpublished  
REFERENCE 2 (bases 1 to 2686)  
AUTHORS Wu, G. and Dorn, G.W. II.  
JOURNAL Direct Submission  
Submitted (19-JAN-2000) Internal Medicine, University of  
Cincinnati, 231 Bethesda Avenue, Cincinnati, OH 45267-0590, USA

FEATURES		Location/Qualifiers	
source	1..2686	/organism="Mus musculus"	
		/db_xref="taxon:10090"	
CDS	29..646	/note="member of RAS oncogene family"	
		/codon_start=1	
		/product="small GTP-binding protein RAB1A"	
		/protein_id="AA03844.1"	
		/db_xref="GI:6969622"	
		/translation="MSSMNPEDYFLKLLIDSGVSKCLLRFADDTYESYISTI GVDFKIRITLEDGKTKIKLIQIMDTAGOEFRITSSYKAGAGIIVYDITDSEFNV KWLQIEDIRASEENVKLLVGNKCDLTKRVVDYTTAKERADSLGIPLETSAKNATN VEQSFPMMAAEIKRRMGGAAGAEKSNVKIQSTPYKQSGGCC"	
BASE COUNT	778 a 520 c 563 g 825 t		
ORIGIN			
Alignment Scores:			
Pred. No.:	2.58e-74	Length:	2686
Score:	860.00	Matches:	173
Percent Similarity:	84.39%	Conservative:	0
Best Local Similarity:	84.39%	Mismatches:	0
Query Match:	97.07%	Indels:	32
DB:	10	Gaps:	1
US-09-820-003a-2 (1-173) x AF226873 (1-2686)			
QY	1	MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer	20
DB	29	ATGTCAGCATGATCCGAAATGATATTTATTCAGATTACTTCTGATGGCGATCT	88
QY	21	GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspThrTyrThrGluSerTyr	40
DB	89	GGGTTGGAAAGCTCCGCTTCCTTACGTTTCAGATGATGATACGGAAGCTAC	148
QY	41	IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle	60
DB	149	ATCAGCACAATGGTGGTGGATTCAGATACGACATACAGTAAATGCGAAGACATC	208
QY	61	LysLeuGlnIle-----	64
DB	209	AAGCTACAGATATGGACACAGCAGCCAGAAAGATTGCAACATCCTCCAGTTAA	268
QY	65	-----GluSerPheAsn	68
DB	269	TACAGAGAGCCCATGATCATAGTTGTATGATGTCAGACATGAGAGTCTTCATAT	328
QY	69	AsnValLysGlnTyrPleuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu	88
DB	329	AACGTTAAACAGGCTCGCAGAGATAGATCGCTACGCCAGTAAATGTCCACAACTTG	388
QY	89	LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrAlaLys	108
DB	389	TTGGTAGGGAACAATGTGACCTGACCAACAAAGTACATACACAAAGCAGCAAG	448
QY	109	GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn	128
DB	449	GAATTTCCAAATATCTTGGAAATTCATTTTGGAAACAGTCTTAAGAACGCAACCAAT	508
QY	129	ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyValA	148
DB	509	GTAACAACAGCTTTATCATCATGACATGCGACGTCGAGATTAAGGAATGGCGTCGAGCT	568
QY	149	ThrAlaGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer	168
DB	569	ACAGCTGTGTGGCCGAGAAATCCAAATGTTAAATCCAGAGCACTCCAGTCAACAGATCA	628
QY	169	GlyLysGlyCysCys 173	
DB	629	GGTGGAGCTGCTGC 643	
RESULT 9			
RATRASA			
LOCUS			
RATRASA 618 bp mRNA linear ROD 27-APR-1993			
DEFINITION			
Rat ras-related protein mRNA, clone NTRAB1R.			
ACCESSION			
J02998			
VERSION			
J02998.1 GI:206552			
KEYWORDS			
ras oncogene.			
SOURCE			
Rat brain, cDNA to mRNA, (library of D Anderson), clone NTRAB1R.			
ORGANISM			
Rattus norvegicus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
1 (bases 1 to 618)			
Touchot, N., Chardin, P. and Tavillan, A.			
Four additional members of the ras gene superfamily isolated by an			
oligonucleotide strategy: molecular cloning of vpr-related cDNAs			
from a rat brain library			
Proc. Natl. Acad. Sci. U.S.A. 84 (23), 8210-8214 (1987)			
JOURNAL			
MEDLINE			
PUBMED			
3317403			
Draft entry and computer-readable sequence for [1] kindly provided			
by N.Touchot, 01-OCT-1987.			
FEATURES			
source			
1..618			
/organism="Rattus norvegicus"			
/db_xref="taxon:10116"			
1..618			
/note="ras protein"			
/codon_start=1			
/protein_id="AAA42006.1"			
/db_xref="GI:206553"			
/translation="MSSMNPEDYFLKLLIDSGVSKCLLRFADDTYESYISTI GVDFKIRITLEDGKTKIKLIQIMDTAGOEFRITSSYKAGAGIIVYDITDSEFNV KWLQIEDIRASEENVKLLVGNKCDLTKRVVDYTTAKERADSLGIPLETSAKNEN VEQSFPMMAAEIKRRMGGAAGAEKSNVKIQSTPYKQSGGCC"			
BASE COUNT			
194 a 124 c 160 g 140 t			
ORIGIN			
unreported.			
Alignment Scores:			
Pred. No.:	5.34e-74	Length:	618
Score:	849.00	Matches:	171
Percent Similarity:	83.41%	Conservative:	0
Best Local Similarity:	83.41%	Mismatches:	2
Query Match:	95.82%	Indels:	32
DB:	10	Gaps:	1
US-09-820-003a-2 (1-173) x RATRASA (1-618)			
QY	1	MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer	20
DB	1	ATGTCAGCATGATCCGAAATGATATTTATTCAGATTACTTCTGATGGCGATCT	60
QY	21	GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspThrTyrThrGluSerTyr	40
DB	61	GGGTTGGAAAGCTTCCTCTCTAGGTTTCAGATGATGATACGGAAGCTAC	120
QY	41	IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle	60
DB	121	ATCAGCACAATGGTGGTGGATTCAGATACGACATACAGTAAATGCGAAGACATC	180
QY	61	LysLeuGlnIle-----	64
DB	181	AAGCTTACAGATATGGACACAGCAGCCAGAAAGCTTGCACAAATCCTCCAGTTAT	240
QY	65	-----GluSerPheAsn	68
DB	241	TACAGAGAGCCCATGATCATAGTTGTATGATGTCAGACACGAGAGTCTTCATAT	300
QY	69	AsnValLysGlnTyrPleuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu	88
DB	301	AACGTTAAACAGTGGCTCGCAGAGATAGATCGCTACGCCAGTAAATGTCCACAACTTG	360
QY	89	LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrAlaLys	108
DB	361	TTGGTAGGGAACAATGTGACCTGACCAACAAAGTACATACACAAAGCAGCAAG	420

QY 109 GluPheAlaaspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
 |||||  
 Db 421 GAATTCGACATTCCTCGAATTCATTTTGGAAACAGCTGCTGAAGAACGAAGAAT 480  
 QY 129 ValGluInserPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
 |||||  
 Db 481 GTAGAACAGCTCTTCATGACACATGCGAGATTAATAAAGCCGATGGTCCCTGGAGCA 540  
 QY 149 ThrAlaGlyAlaGlyAlaGlyLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 |||||  
 Db 541 ACAGCTGAGAGTGGGAGAACTCCATGTTAAATCCAGAGCTCCAGTCAAGACAGTCA 600  
 QY 169 GlyGlyGlyLysCys 173  
 |||||  
 Db 601 GGTGAGAGCTGCTGC 615  
 RESULT 10  
 DYGORABI 841 bp mRNA linear VRT 28-APR-1993  
 LOCUS Discopyge ommata GTP-binding protein (o-rab1) mRNA, complete cds.  
 DEFINITION M38393.1 GI:213122  
 ACCESSION  
 VERSION  
 KEYWORDS GTP-binding protein.  
 SOURCE D.ommata electric lobe electromotor neuron, cDNA to mRNA.  
 ORGANISM Discopyge ommata  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 Elasmobranchii; Squalea; Hypnosqualea; Pristioraja; Batoidae;  
 Torpediniformes; Narcinoidel; Narcinidae; Discopyge.  
 1 (bases 1 to 841)  
 Ngsee,J.K., Elferink,L.A. and Scheller,R.H.  
 A family of ras-like GTP-binding proteins expressed in electromotor  
 neurons  
 Unpublished (1990)  
 JOURNAL Draft entry and computer-readable sequence for [1] kindly submitted  
 COMMENT by J.K.Ngsee, 10-SEP-1990.  
 Dept of Biological Sciences  
 Heriain Lab  
 Stanford University  
 Stanford, CA 94305-5020  
 Draft entry and computer-readable sequence for [1] kindly submitted  
 by J.K.Ngsee, 10-SEP-1990.  
 Dept of Biological Sciences  
 Heriain Lab  
 Stanford University  
 Stanford, CA 94305-5020.  
 Location/Qualifiers  
 1..841  
 /organism="Discopyge ommata"  
 /db\_xref="taxon:7785"  
 /cell\_type="electromotor neuron"  
 /tissue\_type="electric lobe"  
 1..841  
 /gene="o-rab1"  
 79..687  
 /gene="o-rab1"  
 /product="GTP-binding protein"  
 /codon\_start=1  
 /protein\_id="AAA49234.1"  
 /db\_xref="GI:213123"  
 /translation="MNPEYDYLEKLLIGDSGVKSCLLRFADDTFTSYSTISIVD  
 FKIRITELDGKTIKLIQIMDTAGQERFTITSSYRAHGIIIVYDVTQDSEFNVRQW  
 LOEIDRYASENVKILVGNKCDLTKRVVDTYAKREFADSLIGIPLETSAKNATNVEQ  
 AFMTMAEIKIKRMGPATSGSGSKSNVNIOSTPVKSGGGCC"  
 BASE COUNT 269 a 154 c 194 g 224 t  
 ORIGIN

DB: 5 Gaps: 1  
 US-09-820-003a-2 (1-173) x DYGORABI (1-841)  
 QY 1 MetSerSerMetAsnProGluIuTyAspTyLeuPheLysLeuLeuIleGlyAspSer 20  
 |||||  
 Db 70 ATGCAACCATGATGCCGAAATAGACTATTATTCAAGTCGCTGATGGAGATCA 129  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyThrGluSerTy 40  
 |||||  
 Db 130 GGTGTGAAAATTTCTCTCTCTTACATTCACATGATACATATACAGAAAGTTAT 169  
 QY 41 IleSerThrIleGlyAlaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
 |||||  
 Db 190 ATCAGTCAATTTGGTGTGATTTTAAATCAGAACATAGATGATTAGACGCAAGACCATC 249  
 QY 61 LysLeuGlnIle----- 64  
 Db 250 AAACCTCAAAATCTGGACAGCGGTGTGACAGAGCGTTTCGAACAAATCACGTCAGTTAC 309  
 QY 65 -----GluSerPheAsn 68  
 Db 310 TACAGAGGGGCACATGCGCATCATAGTTGTATGATGTACAGACCAAGAGTGTATTAC 369  
 QY 69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyAlaSerGluAsnValAsnLysLeu 88  
 |||||  
 Db 370 AATGTAACCAATGCTTCAGGAATATAGATCGTATGCCAGTGAATAATGTTAACAACTTA 429  
 QY 89 LeuValLysAsnLysCysAspLeuThrThrLysLysValValAspTyThrThrAlaLys 108  
 |||||  
 Db 430 CTGGTGGAAACAATATGATCTGCAACAACAAGAGGTGGTGTATTAACAACGCAAG 489  
 QY 109 GluPheAlaaspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
 |||||  
 Db 490 GAATTCGACACTCCTTGGATCCATTTTGGAAACAGCTGCAAGATGCACAAAT 549  
 QY 129 ValGluInserPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
 |||||  
 Db 550 GTAGAACAGCTTCATGACATGCTGAGATTAATAAAGCAATGGGTCCCTGGAGCT 609  
 QY 149 ThrAlaGlyAlaGlyLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 |||||  
 Db 610 ACATCTGCTGCTCTGTAAGAAATCAATGTAAATCCAAAGCAGCCAGTAGAGTCTCT 669  
 QY 169 GlyGlyGlyLysCys 173  
 |||||  
 Db 670 GGTGAGAGCTGCTGC 684  
 RESULT 11  
 AR070365  
 LOCUS AR070365 925 bp DNA linear PAT 18-FEB-2000  
 DEFINITION Sequence 4 from patent US 5892012.  
 ACCESSION AR070365  
 VERSION AR070365.1 GI:7221253  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unclassified.  
 REFERENCE 1 (bases 1 to 925)  
 AUTHORS Hillman,J.L., Lal,P., Corley,N.C. and Shah,P.  
 TITLE Rab Proteins  
 JOURNAL Patent: US 5892012-A 4 06-APR-1999;  
 FEATURES location/Qualifiers  
 1..925  
 /organism="unknown"  
 BASE COUNT 198 a 263 c 288 g 176 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5.92e-66 Length: 925  
 Score: 768.50 Matches: 156  
 Percent Similarity: 80.39% Conservative: 8  
 Best Local Similarity: 76.47% Mismatches: 7  
 Best Match:

Query Match:	86.74%	Indels:	33
DB:	6	Gaps:	2

US-09-820-003A-2 (1-173) x AR070365 (1-925)

```

Oy      2 SerSerMetAsnProGluTyrAspTyrIleuPheLysLeuLeuIleGlyAspSergLy  21
         : ::::: ::::::::::::::::::::
Db      60 GCGGCATGAACCCCAATAGACTACTCTGTTTAAAGCTGTTTTATTATGGCGACTCAGCC  119
Oy     22 ValGlyLysSerCysLeuLeuLeuArgPheAlaAspThrTyrThrGlySerTyrIle  41
         : ::::::::::::::::::::
Db     120 GGGGCAAGTCATGACCCTCTCTCCGCGTTTGCTGTATACACGTCACACAGAGAATCATAC  179
Oy     42 SerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys  61
         : ::::::::::::::::::::
Db     180 ACCACCACTGGGGGTGACTTCAGATGCCAACATCGACTGAGTGAGCAAACTATCAA  239
Oy     62 LeuGlnIle----- 64
         : ::::::::::::::::::::
Db     240 CTTTCAGATCTGGGACACAGCGGCGCAGAGGTTCCGACCACTTCACACTACTAC  299
Oy     65 ----- GluSerPheAsnSn 69
         : ::::::::::::::::::::
Db     300 CGGGGGCTCATGCATCATCGTGTATGACGTCACATGACAGAGAAATCTCAGCCAA  359
Oy     70 ValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGlyAsnValAsnLysLeuLeu  89
         : ::::::::::::::::::::
Db     360 GTGAAGCAGTGGCTCGAGNAATTAGACGCTATGCCGTCACGAGAAACATCATAGCTCTG  419
Oy     90 ValGlyAsnLysGlyAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGlu  109
         : ::::::::::::::::::::
Db     420 GTGGGCAACAAGAGGAGCACTCACACAAAGGTGTGGACAAACACCAACGCAAGAG  479
Oy    110 PheAlaAspSerIleGlyIleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnVal  129
         : ::::::::::::::::::::
Db     480 TTTCGACAGCTCTGGGCACTCCCTTTCTGGAGACAGGCGCCAAAGATCCACCAATGTG  539
Oy    130 GluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr  149
         : ::::::::::::::::::::
Db     540 GAGCAGGGGTTCATGACCATGGCTGTAATCAAAAAGCGGATGGGCGCTGAGACGCC  599
Oy    150 AlaGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSergLy  169
         : ::::::::::::::::::::
Db     600 TCTGGGGGCG---GACGGGCCCAATCTCAAGATCGACAGACACCCCTGTAAAGCCGCGTGGC  656
Oy    170 GlyLysCysCys 173
         : ::::::::::::::::::::
Db     657 GGTGCTCTGTTCG 668

RESULT 12
AX086155          1985 bp   DNA           linear   PAT 09-MAR-2001
LOCUS              AX086155
DEFINITION        Sequence 107 from Patent WO0112659.
ACCESSION         AX086155
VERSION           AX086155.1 GI:13275888
KEYWORDS
ORGANISM          human.
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS           Wiemann,S
TITLE             Human dnt sequences
JOURNAL           Patient: NO 0112659-A;107 22-FEB-2001;
                  German Human Genome Project (DG)
FEATURES
SOURCE            Location/Qualifiers
                  1..1985
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
BASE COUNT       432 a      607 c      349 g      397 t
ORIGIN
Alignment Scores:

```

Pred. No.:	1.45e-65	Length:	1985
Score:	768.50	Matches:	156
Percent Similarity:	80.39%	Conservative:	7
Best Local Similarity:	76.47%	Mismatches:	8
Query Match:	86.74%	Indels:	33
DB:	6	Gaps:	2

US-09-820-003a-2 (1-173) x AX086155 (1-1985)

QY	2	SerSerMetCaaProGluTyrAspPyrLeuPheLysLeuLeuLeuIleGlyAspSerGly	21
DB	42	GGCGGCATGGAACCCCGAATATGACTACCTGTTAAAGCTGCTTTATGATGGACACTAGGC	101
QY	22	ValGlySerSerCysLeuLeuLeuLeuAArgPheAlaAspAspThrTyrThrGluSerTyrIle	41
DB	102	GGGGCAAGCAATGAGCCTCTCTCGCGTTTGCTGATACACCTACACAGAGACCTACTATC	161
QY	42	SerThrIleLeuValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys	61
DB	162	AGCACATCGGGGTGGACTTCAGATCCGACACCATCGACGTGATGGCAAAACTATCAA	221
QY	62	LeuGluIle-----	64
DB	222	CTTTCAGATCTGGGACACAGCGGGCCAGAGAGGTTCCGACCATCTCCACTATAC	281
QY	65	-----GluSerPheAsnAsn	69
DB	282	CGGGGGCTCATGCATCATCTGTGCTATGACTCATGACACAGAAATCTTACGCCAAC	341
QY	70	ValLysThrPheLeuGluIleAlaAspArgTyrThrAlaSerGlyAsnValAsnLysLeuLeu	89
DB	342	GTGAGAGCATGGGTGGAGAGATATGACGCTATCCAGCAGACAGATCATTAAGCTCTG	401
QY	90	ValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGlu	109
DB	402	GGGGCAACAGAGGACCTTCACACCAAGAGGTGTGGACACACACCGACAGCAAGAG	461
QY	110	PheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal	129
DB	462	TTTGAACACTCTGTGGCATCCCTCTTGGAGACGAGGCCAACAATGCGCAAAATCTC	521
QY	130	GluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetCylProGlyAlaThr	149
DB	522	GAGCAGGCGTTCATGACCATGGCTGTGAATAACAAACGGATGGGGCTGTGACACC	581
QY	150	AlaGlyIleValGluLysSerAsnValLysIleLeuInsThrProValLysGlnSerGly	169
DB	582	TTGTGGGGGCG---GAGCGGGCCCATCTCAAGATCGACGACACCCCTGTAAAGCCGGCTGC	638
QY	170	GlyGlyCysCys	173
DB	639	GGTGGCTGTTC	650

RESULT 13  
HSMB01608  
LOCUS  
DEFINITION  
complete cds.  
ACCESSION  
version  
KEYWORDS  
SOURCE  
ORGANISM

human.  
Homo sapiens  
Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.  
1 (bases 1 to 1985)  
Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and  
Wiemann,S.  
Direct Submission  
Submitted (12-MAR-2002) MIMF, Am Klopferspitz 18a, D-82152  
Martinsried, GERMANY  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;

sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project.  
This clone (DKFZp564i172) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonezrpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

## FEATURES

## source

1. 1985  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="11q13"  
/clone="DKFZp564i172"  
/tissue\_type="brain"  
/clone\_id="564 (synonym: hfbz2). Vector pAMP1; host  
X1-2b1ue; sites NotI + SalI"  
/dev\_stage="fetal"  
1. 1985  
/gene="DKFZp564i172"  
48. 653  
/gene="DKFZp564i172"  
/note="strong similarity to Rab1"  
/product="hypothetical protein"  
/protein\_id="CA86570.1"  
/db\_xref="GI:12052796"  
/db\_xref="SPTREMBL:Q9H0U4"  
/translation="MNPEYDYLFRKLLIGDSGVKSCLLRFADDTYTESYISTIGVD  
FKIRTEIDSGTKIKIOWDPAOGERPRTTSSYRGAHGIIVYDVTDOSYANVKOM  
LQETIDRASENVKLVGNKSDLTFTKKYVDNTAKKEFADSLGTFLETSKKNATNVEQ  
AFMTMAEIKRRKPGASGGERPNLKIDSTPVKASGSGC"  
1860. 1865  
/gene="DKFZp564i172"  
1901  
polya\_site  
/gene="DKFZp564i172"  
BASE COUNT 432 a 607 c 549 g 397 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.46e-65 Length: 1985  
Score: 768.50 Matches: 156  
Percent Similarity: 80.39% Conservative: 8  
Best Local Similarity: 76.47% Mismatches: 7  
Query Match: 86.74% Indels: 33  
DB: Gaps: 2

US-09-820-003a-2 (1-173) x HSM801608 (1-1985)

QY 2 SerSerMetAsnProGluTYrAspTYrLeuPheLysLeuLeuIleGlyAspSerGly 21  
Db 42 GCGCCCATGAAACCCCGCATATGACTCTTAAAGCTGCTTTGATTGGCGACTCAAGC 101  
QY 22 ValGlyLysSerGlyLeuLeuLeuArgPheAlaAspAspThrTYrThrGluSerTYrIle 41  
Db 102 GTGGGCAAGTCATCCTGCTCTCGGGTTGCTGATGACAGTACACAGAGACTACATC 161  
QY 42 SerThrIleGlyValaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61  
Db 162 AGCACCAATCGGGGTGAGACTTCAAGATCGAACAATCGAGCGTGATGGCAAAACATATCAA 221  
QY 62 LeuGlnIle----- 64  
Db 222 CTTCAAGATCTGGACACAGCGGGCCAGGAACGGTTCCGAGACATCCAGCTACTAC 281  
QY 65 -----GluSerPheAsnAsn 69  
Db 282 CCGGGGCGCTCATGCGATCATCGTGCTGTATGACCTCACTGACCAAGAACTCTACGCCAAC 341  
QY 70 ValIleGlnTrpLeuGlnIleValaAspArgTYrAlaSerGluAsnValaLysLeuLeu 89  
Db 342 GTGAAGCAGAGTGGCTGACGAGATGACCGCATGCCAGCGAAGACGTCAATACCTCTG 401  
QY 90 ValGlyAsnLysCysAspLeuThrThryLysValaValaAspTYrThrThryAlaLysGlu 109

Db 402 GTGGGCAACAGACGCGACCTCACCAAGAGGTGGTGGACCAACACACAGCCACAGAG 461  
QY 110 PheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129  
Db 462 TTTCAGACTCTTCGGGATCCCTTCTTGGAGACGACGCCAAGAAATGCCACAAATGTC 521  
QY 130 GluInSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr 149  
Db 522 GAGCAGCGGTCATGACCATGCTGTGTAATCAAAAGCGGATGGGCGCTGGAGACGCC 581  
QY 150 AlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 169  
Db 582 TCTGGGGGCGC---GAGCGCGCCAAATCTCAAGATGACACACACCCCTGTAAAGCGGCTGGC 638  
QY 170 GlyGlyCysCys 173  
Db 639 GGTGGCTGTTC 650

## RESULT 14

RNRAB1B  
LOCUS RAB1B 654 bp mRNA linear ROD 12-SEP-1993  
DEFINITION Rat cDNA for ras-related rab1b protein.  
ACCESSION X13905  
VERSION X13905.1 GI:57005  
KEYWORDS GTP-binding protein; rab1b gene; rab1b protein; ras-related protein.

## SOURCE

ORGANISM Rattus sp.  
Rattus sp.  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

## REFERENCE

1 (Bases 1 to 654)  
AUTHORS Touchot, N.  
TITLE Direct Submission  
JOURNAL Submitted (03-JAN-1989) Touchot N., Inserm U 248, Faculte de  
Medicine Lariboisiere St Louis, 10 Avenue de Verdun, 75010 Paris,  
France  
2 (Bases 1 to 654)  
AUTHORS Vieh, E., Touchot, N., Zahraoui, A. and Tavittian, A.  
TITLE Nucleotide sequence of a rat cDNA: rab1b, encoding a rab1-YPT  
related protein  
JOURNAL Nucleic Acids Res. 17 (4), 1770 (1989)  
MEDLINE 89160341  
PUBMED 2493636

COMMENT Data kindly reviewed (13-MAR-1989) by Touchot A.  
FEATURES Location/Qualifiers

## source

1. 654  
/organism="Rattus sp."  
/db\_xref="taxon:10118"  
/tissue\_type="brain"  
12. 617  
/note="rab1b protein (AA 1 - 201)"  
/codon\_start=1  
/protein\_id="CAA32105.1"  
/db\_xref="GI:57006"  
/db\_xref="SWISS-PROT:P10536"  
/translation="MNPEYDYLFRKLLIGDSGVKSCLLRFADDTYTESYISTIGVD  
FKIRTEIDSGTKIKIOWDPAOGERPRTTSSYRGAHGIIVYDVTDOSYANVKOM  
LQETIDRASENVKLVGNKSDLTFTKKYVDNTAKKEFADSLGTFLETSKKNATNVEQ  
AFMTMAEIKRRKPGASGGERPNLKIDSTPVKASGSGC"

BASE COUNT 174 a 156 c 184 g 140 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.51e-65 Length: 654  
Score: 762.50 Matches: 154  
Percent Similarity: 79.90% Conservative: 9  
Best Local Similarity: 75.49% Mismatches: 8  
Query Match: 86.06% Indels: 33  
DB: Gaps: 2

US-09-820-003a-2 (1-173) x RNRAB1B (1-654)







Mon Nov 18 08:42:03 2002

us-09-820-003a-2.rge

Page 13

Dh 160554 GACAGGCATTCATGACATATGGCTCAGAGAATCATAAAAACCGATGTGGGGCCAGAGACAGCA 160435  
QY 150 AIAIGVYLVAIAGLILySerrsnValIIIEgInserThprOvalLYGIInserLy 169  
:::||||| ::|||::| ||||| ||||| |||||  
Db 160494 TCTGGGGT---GAACGCCCAACTGTAAGATGACACACACTCTGTGAAATCTGCTAGT 160438  
QY 170 GLVGLVYCVCY 173  
||||| ||||| ||||| ||||| |||||  
Db 160437 GGTCGCTGCTGC 160426

Search completed: November 17, 2002, 13:58:04  
Job time : 2672 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 17, 2002, 13:04:33 : Search time 254 Seconds  
(without alignments)

1533.841 Million cell updates/sec

Title: US-09-820-003a-2

Perfect score: 886

Sequence: 1 MSNNPEYDYLFKLLIGDS.....EKSNNKIOSTPVKSGGCGC 173

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-KODEL-frame\_p2n.model -DEV-xlh  
-Q-cgn2.1/USPFO.spool/US0982003/rnatc.11112002.134616.3164/app.query.fasta.1.327  
-DB-N.Geneseq.101002 -OFM-fastap -SUFFIX-rrng -MINMATCH-0.1 -LOOPT-0  
-LOOPT-0 -UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human40.cdi  
-LIST-43 -DOCALLIGN-200 -THR-SCORE-pct -THR\_MAX-100 -THR\_MIN-0 -ALIGN-15  
-MODE-LOCAL -OUTFM-pfo -NORM-ext -HEADSITE-500 -MINLEN-0 -MALEN-2000000000  
-USER-US0982003.ecgn.1.125-errnat.11112002.134616.3164 -NCU-6 -ICPU-3  
-NO\_XLUPX -NO\_MMAP -LARGQUERY -NEG\_SCORES-0 -WAIT -LONLOG -DEV-TIMEOUT-120  
-WARN\_TIMEOUT-30 -THREDS-1 -XGAPOP-10 -XGAPEXT-0.5 -Fgapop-6 -Fgapext-7  
-Ygapop-10 -Ygapext-0.5 -DELOP-6 -DELEXT-7

Database :  
1: N\_Geneseq.101002.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860	97.1	723	24	ABK83978 Human cDNA differe
2	860	97.1	2538	21	AAC60009 Human secreted pro
3	768.5	86.7	925	20	AAK27232 Human Rab protein,
4	768.5	86.7	939	21	AAF21661 Human breast and o
5	768.5	86.7	1898	22	AAK49244 Human encoding nove
6	768.5	86.7	1944	22	AAI93456 Human polynucleoti
7	733.5	82.8	1965	22	AAK34668 Human DNA for a no
8	683	77.1	1791	23	AAK5983 Arabidopsis thaliana
9	657.5	74.2	1202	21	AAK14053 Arabidopsis thaliana
10	639	72.1	959	21	AAK3987 Arabidopsis thaliana
11	633	71.4	932	21	AAK35200 Arabidopsis thaliana
12	624	70.4	1193	23	AAK3860 DNA encoding novel
13	619.5	69.9	777	21	AAK2684 Arabidopsis thaliana
14	619.5	69.9	881	24	ABN8761 Arabidopsis thaliana
15	612	69.1	666	21	AAK2764 Arabidopsis thaliana
16	612	69.1	932	21	AAK48274 Arabidopsis thaliana
17	597	67.4	607	22	AAK87926 Arabidopsis thaliana
18	580	65.5	584	21	AAK4951 Arabidopsis thaliana
19	570	64.3	1380	21	AAK7578 Arabidopsis thaliana
20	529	59.3	5019	23	ABK0582 Arabidopsis thaliana
21	460	51.9	743	21	AAK14936 Arabidopsis thaliana
22	442	50.1	911	21	AAK38429 Arabidopsis thaliana
23	442	49.9	1101	21	AAK34080 Arabidopsis thaliana
24	440.5	49.7	1203	21	AAK33685 Arabidopsis thaliana
25	431	48.6	1129	21	AAK43682 Arabidopsis thaliana
26	429.5	48.4	1035	21	AAK9208 Arabidopsis thaliana
27	428.5	48.4	1035	21	AAK51491 Arabidopsis thaliana
28	428.5	48.4	1035	21	AAK3437 Arabidopsis thaliana
29	419	47.3	1966	22	AAK09160 Human rab8 homolog
30	417	47.1	2411	24	ABO54182 Human ovarian anti
31	414	46.7	1161	22	AAK75182 Nucleotide sequenc
32	414	46.7	2497	22	AAK60878 Human cancer agent
33	414	46.7	2497	22	AAK60878 Human cancer agent
34	414	46.7	2497	22	AAK60878 Human cancer agent
35	414	46.7	2497	22	AAK60878 Human cancer agent
36	414	46.7	2497	22	AAK60878 Human cancer agent
37	414	46.7	2497	23	ABV25781 Human prostate exp
38	414	46.7	2497	23	AAH3912 Human prostate exp
39	414	46.7	3077	22	AAH4301 Human cDNA sequenc
40	413	46.6	866	22	AAH4301 Human cDNA sequenc
41	408	46.0	2247	23	ABL29661 Drosophila melanog
42	405	45.7	492	21	AAK38136 Zea mays DNA fragm
43	405	45.7	559	21	AAK44411 Arabidopsis thaliana
44	403.5	45.5	716	21	AAA40104 Human Rab10 cDNA
45	403.5	45.5	861	21	AAA40108 Human Rab10 cDNA

## ALIGNMENTS

RESULT 1  
ABK83978  
ABK83978 standard; cDNA: 723 BP.  
ABK83978;  
14-AUG-2002 (first entry)  
Human cDNA differentially expressed in granulocytic cells #549.  
Human; ss: granulocytic cell; DNA chip: bacterial infection;  
viral infection; parasitic infection; protozoal infection;  
fungal infection; sterile inflammatory disease; psoriasis;  
rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
cardiac reperfusion injury; renal reperfusion injury; AIDS;  
adult respiratory distress syndrome; inflammatory bowel disease;  
Crohn's disease; ulcerative colitis; periodontal disease;  
granulocyte activation; chronic inflammation; allergy.



```

XX  Rosen CA, Ruben SM, Komatsoulis G;
PI  WPI: 2000-594639/56.
XX  P-PSDB: AAB34816.
DR  P-PSDB: AAB34816.
XX  Fifty nucleic acid molecules encoding human secreted proteins, useful
PT  in the prevention, treatment and diagnosis of cancer, immune disorders,
PT  cardiovascular disorders and neurological diseases -
XX  Claim 1: Page 371, 425pp; English.
PS
CC  The polynucleotide sequences given in AAC59966 to AAC60015 encode the
CC  human secreted proteins given in AAB34773 to AAB34822. AAB34823 to
CC  AAB34852 represent human secreted polypeptide sequences and proteins
CC  homologous to them, which are given in the exemplification of the present
CC  invention. Human secreted proteins have activities based on the tissues
CC  and cells the genes are expressed in. Examples of activities include:
CC  cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
CC  antiatheric; hepatotropic; antibacterial; antifungal; antiparasitic; and
CC  cardiant. The polynucleotides and polypeptides are useful for
CC  preventing, treating or ameliorating a medical condition in e.g. humans,
CC  mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC  polypeptides can also be used as a food additive or preservative to
CC  increase or decrease storage capabilities. The polynucleotides are
CC  useful for chromosome identification. They are also useful as probes for
CC  diagnosing a disorder related to the female reproductive system,
CC  particularly breast and/or ovary cancer. They are also useful in the gene
CC  therapy of breast and ovarian cancer. Nucleic acids, protein, antibodies,
CC  agonists and antagonists from the present invention are useful in the
CC  diagnosis, treatment and prevention of cancer, immune disorders,
CC  cardiovascular disorders, wound healing, neurological diseases and
CC  infectious disease. AAC59957 to AAC59965 and AAB34772 represents sequence
CC  used in the exemplification of the present invention.
XX
XX  Sequence 2528 BP; 772 A; 428 C; 545 G; 781 T; 2 other:
SQ
Alignment Scores:
Pred. No.: 2,24e-88 Length: 2528
Score: 860.00 Matches: 173
Percent Similarity: 84.39% Conservative: 0
Best Local Similarity: 84.39% Mismatches: 0
Query Match: 97.07% Indels: 32
DB: Gaps: 1
US-09-820-003a-2 (1-173) x AAC60009 (1-2528)
OY  1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20
DB  208 ATGTCACGATGATGATCGAATATGATTTATTCAGTCTCTGATTCGCGACTCA 267
OY  21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrcyluSerTyr 40
DB  268 GGGGTGGAAAGCTTGGCTTCTTACGCTTTCGACATGATCATATACACAAAGCTAC 327
OY  41 IleSerThrIleGlyValAspPheLysIleArgPheIleGluLeuAspLysThrIle 60
DB  328 ATCCAGCATATTCGTGTGATTTCAAAATTAAGACTATAGAGTTAAGCGGAAACAATC 367
OY  61 LysLeuGluIle----- 64
DB  388 AAGCTTAATATATGAGACACACAGCGCCAGGAAATTTGCAACAATCCTCCATTTAT 447
OY  65 -----GluserPheasn 68
DB  448 TACAGAGACCCATGCGATATAGTTGTGTATGATGACAGATCAGAGGCTCTCAAT 507
OY  69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88
DB  508 AATGTTAAACAGGCTGCTCAGGAAATGATCGTTCATGCGACGTAAATCATCAAAATG 567
OY  89 LeuValGlyLysLysCysAspLeuThrLysLysValValAspTyrThrAlaLys 108

```

```

DB  568 TTGCTAGGCAACAAATCTATCTGACCAAAAGATGATGACTACACAAAGCGAG 627
OY  109 GluPheAlaAspSerLeuGluIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128
DB  628 GAATTTGCTGATCTGCTGGATTCGGATTTTGGAAACGAGTGCATGAAGATGACAGAT 687
OY  129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148
DB  688 GTAGACACATCTTTCATGACGATGACGACGTGACATTTAAACCAATGAGGCTCCGAGACA 747
OY  149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168
DB  748 ACAGCTGGTGGTGGTGGAGATCCATGTTAAATTCACAGCAGCTCCATCAGACATCA 807
OY  169 GlyGlyGlyCysCys 173
DB  808 GGTGAGGCTGCTGTC 822
RESULT 3
AAX27232
ID  AAX27232 standard; DNA; 925 BP.
XX
XX  AAX27232;
AC
XX  28-MAY-1999 (first entry)
DE
XX  Human Rab protein, RABP-2, coding sequence.
KW
KW  Rab protein; RABP-1; RABP-2; RABP-3; human; vesicle trafficking; cancer;
KW  cell differentiation; apoptosis; immunodeficiency; cell proliferation;
KW  neurodegenerative disease; myelodysplastic syndrome; wasting disease;
KW  toxin-induced disease; infection; genetic defect; diagnosis; therapy; ss.
OS  Homo sapiens.
XX
XX  MO9909182-A2.
XX
XX  25-FEB-1999.
XX
XX  17-AUG-1998; 98MO-US16983.
XX
XX  21-AUG-1997; 97US-0916901.
XX
XX  (INCY-) INCYTE PHARM INC.
PA
XX  Corley NC, Hillman JL, Lal P, Shah P;
PI  WPI: 1999-181042/15.
DR  P-PSDB: AAY00919.
XX
XX  New purified human Rab proteins - used to develop products for
XX  treating e.g. AIDS, immunodeficiencies, neurodegenerative diseases,
XX  cancer, inflammation or autoimmune diseases.
PS  Claim 29: Fig 2, 94pp; English.
XX
XX  This sequence encodes a human Rab protein of the invention, designated
XX  RABP-2. The Rab proteins, RABP-1, RABP-2 and RABP-3 are involved in
XX  vesicle trafficking, cell function, and cell differentiation. The RABP
XX  polypeptides, DNAs and agonists can be used to prevent or treat a
XX  disorder associated with an increase in apoptosis, e.g. infectious or
XX  genetic immunodeficiencies, neurodegenerative diseases such as
XX  Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis,
XX  retinitis pigmentosa, and cerebellar degeneration, myelodysplastic
XX  syndromes such as aplastic anemia, ischemic injuries such as myocardial
XX  infarction, stroke, and reperfusion injury, toxin-induced diseases such
XX  as alcohol-induced liver damage, cirrhosis, and lathyrism, wasting
XX  diseases such as cachexia, viral infections, and osteoporosis. They can
XX  also be used to stimulate cell proliferation for use in transplantation
XX  or to produce cells to fight an infection or a cancer or to correct a
XX  genetic defect in a disease such as sickle cell beta thalassemia, cystic
XX  fibrosis or Huntington's chorea. Antagonists can be used to prevent or

```

CC treat a disorder associated with cell proliferation e.g. cancers or  
 CC inflammation, e.g. Addison's disease, adult respiratory distress  
 CC syndrome, allergies, asthma, atherosclerosis, bronchitis, cholelithiasis,  
 CC Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis,  
 CC diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis,  
 CC gout, Graves' disease, hyperostosis, irritable bowel syndrome, lupus  
 CC erythematous, multiple sclerosis, myasthenia gravis, inflammation,  
 CC osteoarthritis, osteoporosis, pancreatitis, polyomyelitis, rheumatoid  
 CC arthritis, scleroderma, Sjogren's syndrome, and autoimmune thyroiditis,  
 CC complications of cancer, haemodialysis, extracorporeal circulation,  
 CC infections and trauma.  
 XX  
 SQ Sequence 925 BP; 198 A; 263 C; 288 G; 176 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,78e-78 Length: 925  
 Score: 768.50 Matches: 156  
 Percent Similarity: 80.39% Conservative: 8  
 Best Local Similarity: 76.47% Mismatches: 7  
 Query Match: 86.74% Indels: 33  
 DB: 20 Gaps: 2

US-09-820-003a-2 (1-173) x AAF21661 (1-925)

QY 2 SerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSergly 21  
 Db 60 GCGCCCATGAACCCCGAATATGACTGTTTAAGCTGCTTTGATTGGCGACTGAGC 119  
 QY 22 ValGlyLysSeryLeuLeuLeuArgPheAlaAspSeryTyrThrGluSeryTyrIle 41  
 Db 120 GTGGGCAAGTATCGCTGCTCTCGGTGCTGATGACAGTACAGAGAGCTACATC 179  
 QY 42 SerThrIleGlyValAlaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61  
 Db 180 AGCAGCATCGGGGGGAGACTTCAATCATCGAAGCATCGAGCTGATGGCAAAACATCAA 239  
 QY 62 LeuGlnIle----- 64  
 Db 240 CTTGAGATCTGGGACAGAGCGGGCCAGGAAAGCTTCGAGCATCTCCAGCTACTAC 299  
 QY 65 -----GluSerPheAsnAsn 69  
 Db 300 CGGGGGGCTATGCGCATCATCTGCTGTGATGACCTACTGACCGAGAACTCTACGCCAAC 359  
 QY 70 ValLysGlnTyrLeuGlnGlnIleAspArgTyrAlaSeryGluAsnValAlaLysLeuLeu 89  
 Db 360 GTGAAGCAGTGGCTGACGAGATGACCGGTATGCCAGCGAAGCTCAATTAAGCTCTCTG 419  
 QY 90 ValGlyAsnLysCysAspLeuThrThrLysLysValValAlaAspTyrThrAlaLysGlu 109  
 Db 420 GTGGGCAACAAAGAGCGCTTCCACCAAGAGTGGTGGCAACACCCAGCAAGAGAG 479  
 QY 110 PheAlaAspSerLeuGlnIleProPheLeuGlnThrSerAlaLysAsnAlaThrAsnVal 129  
 Db 480 TTTCACACATCTCTGGGCAATCCCTTCTTGGAGACGAGCGCCAAAGATGCAACAAATGTC 539  
 QY 130 GluGlnSerPheMetThrMetAlaAlaGlnIleLysLysArgMetGlyProGlyAlaThr 149  
 Db 540 GAGCAGCGCTCATGACCATGAGCTGCTGAATCAAAAAGGAGGGGCGCTGGAGCAGCC 599  
 QY 150 AlaGlyLysAlaGlnLysSerAsnValLysIleGlnSerThrProValLysGlnSergly 169  
 Db 600 TCTGTGGGGGCT--GAGCGCCCAATCTCAAGATCGACAGCACCCTGTAAAGCCGCGTGC 656  
 QY 170 GlyGlyCysCys 173  
 Db 657 GGtGGCTGTTC 668

RESULT 4  
 AAF21661  
 ID AAF21661 standard; DNA: 939 BP.  
 XX AAF21661;  
 AC  
 4

XX 27-MAR-2001 (first entry)  
 DT Human breast and ovarian cancer associated antigen gene SEQ ID 48.  
 XX  
 DE Human: breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
 KW nototropic; neuropeptide; antiviral; antifungal; hepatotropic;  
 KW antidiabetic; antiparasitic; antiviral; antileukemic; anticonvulsant;  
 KW antidiabetic; antifungal; antiparasitic; cardiac; immune disorder;  
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;  
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KW cardiovascular disorder; wound healing; neurological disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200055173-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000MO-US05881.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI, 2000-611515/58.  
 XX  
 DR P-PSDB; AAB58738.  
 XX

New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention,  
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
 PT disorders and neurological diseases -  
 XX  
 PS Claim 1; Page 515; 1299pp; English.

Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the  
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
 CC nototropic; neuroprotective; antiviral; antileukemic; hepatotropic;  
 CC antidiabetic; antiparasitic; antiviral; antileukemic; anticonvulsant;  
 CC antibacterial; antifungal; antiparasitic and cardiac activity. The  
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
 CC particularly breast and ovarian cancer. The nucleic acid sequences,  
 CC proteins, agonists and antagonists may also be used in the diagnosis,  
 CC prevention and treatment of immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
 CC arthritis and ulcerative colitis; cardiovascular disorders such as  
 CC myocardial ischaemia; wound healing; neurological diseases such as  
 CC cerebral anoxia and epilepsy; and infectious diseases.  
 XX

SQ Sequence 939 BP; 202 A; 264 C; 289 G; 180 T; 4 other;

Alignment Scores:  
 Pred. No.: 1,82e-78 Length: 939  
 Score: 768.50 Matches: 156  
 Percent Similarity: 80.39% Conservative: 8  
 Best Local Similarity: 76.47% Mismatches: 7  
 Query Match: 86.74% Indels: 33  
 DB: 21 Gaps: 2

US-09-820-003a-2 (1-173) x AAF21661 (1-939)

QY 2 SerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSergly 21  
 Db 64 GCGCCCATGAACCCCGAATATGACTGTTTAAGCTGCTTTGATTGGCGACTGAGC 123



D 166 AGCACATCGGGGTGGACTTCAGATCCGAAACCATCGAGCTGGATGCAAAAATATCAAA 225  
QY 62 LeuGlnIle----- 64  
D 226 CTTGAGATCTGGACACAGCGGGCCAGGAGCGTTCGAGACCATCTTCAGCTACTAC 285  
QY 65 -----GlutSerPheAsnAsn 69  
D 286 CGGGGGGCTCATGTCATCATGCTGTGTATGACGTCACTGACCAAGAAATCCTACGCCAAC 345  
QY 70 ValIysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89  
D 346 GTGAAGCAGTGGCTGCAGAGATTGACCGCTATGCCAGCGAAGCAATGAATGCTCTG 405  
QY 90 ValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrAlaLysGlu 109  
D 406 GTGGGCAACAGACGACCTCACCACCAAGAGGTGGTGACCAACACACACCAAGAGAG 465  
QY 110 PheAlaAspSerLeuGlnIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129  
D 466 TTTGCAGACTCTGTGGGATCCCTCTTGTGAGACGACGGCCAGAAATGCCACCAATGTC 525  
QY 130 GluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr 149  
D 526 GAGCAGGCGCTCATGACCATGCTGCTGAATCAAAAAGCGGATGGGCTCGAGACACCC 585  
QY 150 AlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 169  
D 586 TCTGGGGGCG---GAGCGGCCCAATCTCAAGATCGACGACACCCCTGTAAAGCCGGCTGCG 642  
QY 170 GlyGlyCysCys 173  
D 643 GGTGGCTGTTCG 654  
RESULT 6  
AA193456  
ID AA193456 standard; cDNA, 1944 BP.  
AC AA193456;  
XX 06-NOV-2001 (first entry)  
DE Human polynucleotide SEQ ID NO 13516.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukemia;  
KW nervous system disorders; arthritis; inflammation; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001MO-US04927.  
XX  
XX 28-FEB-2000; 2000US-0515126.  
XX  
XX 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
XX  
XX P-PSDB; AA013525.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 1; SEQ ID NO 13516; 1399pp + Sequence Listing; English.  
XX

CC The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX Sequence 1944 BP; 372 A; 618 C; 552 G; 402 T; 0 other;  
SQ  
Alignment Scores:  
Pred. NO.: 4.72e-78 Length: 1944  
Score: 768.50 Matches: 156  
Percent Similarity: 80.39% Conservative: 8  
Best Local Similarity: 76.47% Mismatches: 7  
Query Match: 86.74% Indels: 33  
DB: gaps: 2  
US-09-820-003a-2 (1-173) x AA193456 (1-1944)  
QY 2 SerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGly 21  
D 56 GCGCGCATGAACCCGGAATATGACTACCTGTTAAGCGCTTTGATGGGAGATCAAGCC 115  
QY 22 ValGlyLysSerCysLeuLeuLeuArgPheAlaAspThrTyrThrGluSerTyrIle 41  
D 116 GTGGGCAAGTCATCCTGCTCTCTCGGTTTGTCTGATGACAGTACAGAGAGCTACATC 175  
QY 42 SerThrIleGlyValaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61  
D 176 AGCACCATCGGGGTGGACTTCAGATCGAAGATCGAAGATGAGATGGCAAAATATCAAA 235  
QY 62 LeuGlnIle----- 64  
D 236 CTTGAGATCTGGACACAGCGGGCCAGGAAAGTCCGAGACCATCTCCAGCTACTAC 295  
QY 65 -----GlutSerPheAsnAsn 69  
D 296 CGGGGGGCTCATGTCATCATGCTGTGTATGACGTCACTGACCAAGAAATCCTACGCCAAC 355  
QY 70 ValIysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89  
D 356 GTGAAGCAGTGGCTGCAGAGATTGACCGCTATGCCAGCGAAGCTCAATGAAGCTCTCTG 415  
QY 90 ValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrAlaLysGlu 109  
D 416 GTGGGCAACAGACGACCTCACCACCAAGAGGTGGTGAGCAACACACACCAAGAGAG 475  
QY 110 PheAlaAspSerLeuGlnIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129  
D 476 TTTGCAGACTCTGTGGGATCCCTCTTGTGAGACGAGCGCCAGAAATGCCACCAATGTC 535  
QY 130 GluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr 149  
D 536 GAGCAGGCGCTCATGACCATGCTGCTGAATCAAAAAGCGGATGGGCTCGAGACGCC 595  
QY 150 AlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 169  
D 596 TCTGGGGCG---GAGCGGCCCAATCTCAAGATCGACGACACCCCTGTAAAGCCGGCTGCG 652  
QY 170 GlyGlyCysCys 173  
D 653 GGTGGCTGTTCG 664  
RESULT 7  
AAS34668



ID AAS34668 standard; DNA; 1965 BP.  
XX AAS34668;  
AC  
XX  
XX 17-DEC-2001 (first entry)  
DT  
XX  
XX Human DNA for a novel foetal antigen, SEQ ID No 2092.  
DE  
XX  
XX Human: foetal tissue antigen; ds; antiinflammatory; neuroprotective;  
KW immunomodulator; cardiovascular; cytosolic; nephrotoxic;  
KW cardiovascular; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; breast neoplasm; cancer;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.  
XX  
XX Homo sapiens.  
XX  
XX WO200155312-A2.  
PN  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01321.  
PE  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189876.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 23-AUG-2000; 2000US-0227182.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232387.  
PR 14-SEP-2000; 2000US-0232388.  
PR 14-SEP-2000; 2000US-0232389.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244829.  
PR 08-NOV-2000; 2000US-0244517.  
PR 08-NOV-2000; 2000US-0244574.  
PR 08-NOV-2000; 2000US-0244575.  
PR 08-NOV-2000; 2000US-0244716.  
PR 08-NOV-2000; 2000US-0244717.  
PR 08-NOV-2000; 2000US-0245223.  
PR 08-NOV-2000; 2000US-0245224.  
PR 08-NOV-2000; 2000US-0245225.  
PR 08-NOV-2000; 2000US-0245226.  
PR 08-NOV-2000; 2000US-0245227.  
PR 08-NOV-2000; 2000US-0245228.  
PR 08-NOV-2000; 2000US-0245328.  
PR 08-NOV-2000; 2000US-0245329.  
PR 08-NOV-2000; 2000US-0245332.  
PR 08-NOV-2000; 2000US-0245609.  
PR 08-NOV-2000; 2000US-0245610.  
PR 08-NOV-2000; 2000US-0245611.  
PR 08-NOV-2000; 2000US-0245613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-488782/53.

DR  
XX  
XX  
PT New polynucleotides and polypeptides for diagnosing, treating,  
PT preventing or prognosing e.g. diseases or disorders of the nervous,  
PT musculoskeletal, excretory, gastrointestinal, reproductive, and in  
PT respiratory systems -  
XX  
PS  
XX Disclosure; SEQ ID NO 2092; 642pp; English.

XX  
CC The invention relates to novel nucleic acids encoding novel human foetal  
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. The antibodies to the antigens can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. Numerous  
CC examples of diseases and disorders treated by the nucleic acids and  
CC proteins are given in the specification. The present sequence  
CC is a genomic DNA fragment from a gene encoding a foetal antigen of the

Alignment Scores:

Pred. No.: 4,86e-74 Length: 1965  
Score: 733.50 Matches: 150  
Percent Similarity: 78.92% Conservative: 11  
Best Local Similarity: 73.53% Mismatches: 10  
Query Match: 82.79% Indels: 33  
DB: 22 Gaps: 2

US-09-820-003A-2 (1-173) x AAS34668 (1-1965)

QY 2 SerSetwetsnProgluTyRAspTYrLeuPheLysLeuLeuLeuIleGlyAspSerGly 21

DB 68 GCGGCGCTGGAACCCGCGATGACCTGCTTTAAAGCTGCTTTGATTGGCGAGCTACAGGC 127

QY 22 ValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTrpMgIuserTYrIle 41

DB 128 GTGGGCAAGTACGCTGCTGCGGTTCGATGATACCCCTTACACAGAGAGCTACATC 187

QY 42 SerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61  
|||||  
DB 188 AGCACCATCGGGGTGGACTTCAAGATCCAAACATCAGCTGATGGCAAACTATCAAA 247  
QY 62 LeuGlnIle----- 64  
DB 248 CTTTCAGATCTGGGACACGCGGGGCCAGAAAGGTTCTTGACCATCACTTCACCTACTAC 307  
QY 65 -----GuserPheAsn 69  
DB 308 CCGGGGCTCATGCTTCCCTGCTATATGAGCTCATCAACAGAGATCTTATGCCAAC 367  
QY 70 ValLysGlnTrpLeuGlnGluIleAspArgTYrAlaSerGluAsnValAsnLysLeu 89  
|||||  
DB 368 GTGAAGCAGTGGCTGACGAGATTGACCGCATGCCAGCAAGCTCAATTAAGCTCTG 427  
QY 90 ValGlyAsnLysCysAspLeuThrThyLysValValAspTYrThrAlaLysGlu 109  
|||||  
DB 428 GTGGGCAACAAAGCGCTCACCCACCAAGAGTGGTGGCAACACACGACGCAAGAG 487  
QY 110 PheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129  
|||||  
DB 488 TTTCAGACTCTTTGGGATCCCTTCTTGAGACGAGTGCACAAATGCCACCAATGTC 547  
QY 130 GluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr 149  
|||||  
DB 548 GAGCAGCGCTTCATGACCATGAGCTGCTGAGATCAAAAAGCAGATGGGCGCCGAGCAGCC 607  
QY 150 AlaGlyLysAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 169  
:::||||| |||||  
DB 608 TTCGGGGGCGC---GAGCGGCCCAATCTCAGATGACGACACCCCTGTAAGCGCGTGGC 664  
QY 170 GlyGlyCysCys 173  
|||||  
DB 665 GGTGCTGTTCG 676  
RESULT 8  
ID ABL05983 standard; cDNA; 1791 BP.  
XX ABL05983:  
AC ABL05983:  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12431.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-0509231.  
PF  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PMD, Myers EW;  
XX WPI; 2001-656860/75.  
DR P-PsDB; ABB61880.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
PS Claim 1; SEQ ID NO 12431; 21pp + Sequence Listing; English.



Db 180 TCCATCATGACCCCTGATATCGACTATCTTCAAGCTCCTTCTTATTCGAGATCCGGT 239  
QY 22 ValGlySerCysLeuLeuArgPheAlaAspThrThrGluSerTyrIle 41  
|||||  
Db 240 GTTGGAAATCTTGCTGCTACTCGGTTTGCAAGACACACTACAGAGAGTATATC 299  
QY 42 SerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLys 61  
|||||  
Db 300 TCCACTATGCTGTGATTATTAATCCGAACATCAACTGATGCAAGACAGTGAAG 359  
QY 62 LeuGlnIle----- 64  
|||||  
Db 360 CTTACGATTTGGACACTGCGGCCAGAGCGTTCCGACACCATCAGCTGCTTACTAT 419  
QY 65 -----GluSerPheAsnAsn 69  
|||||  
Db 420 CGAGGTCTCATGTATCTGTGCTGATGATGTTACTGATGATGATGATGATGATGAT 479  
QY 70 ValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89  
|||||  
Db 480 GTGAAGACGTGCTCCAGGAGATCGATCGCTATGCCACTAGCGGTGTCACACAGCTGCT 539  
QY 90 ValGlyAsnLysCysAspIleThrThrLysLysValValAspTyrTrpThrAlaLysGlu 109  
|||||  
Db 540 GTGGGTACACAGAGTACATGGAAGATAAAGGTCTGAGATCACGCTGCCAAAGGAG 599  
QY 110 PheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129  
|||||  
Db 600 TTGCGTATATGCTTGGAATATCACTTCTCGAGACCTCTGCTAAGATGCTCGCAACGTC 659  
QY 130 GluGlnSerPheMetThrMetAlaIleGluIleLysLysArgMetGlyProGlyAlaThr 149  
|||||  
Db 660 GAGCAACACCTTCTTGACAAAGGCAAGGACATCAAGGAGGTATGGGT-----ACC 710  
QY 150 AlAGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly 169  
|||||  
Db 711 GCCACTTCAACACAGCGGACTGTGAGGTGGCCAGGCCAGGGTGTCCAGTCTGGG 770  
QY 170 -----GlyGlyCysCys 173  
|||||  
Db 771 TCCGACGTGTGCTGCTGC 788  
AC 33987;  
AC 33987;  
ID AAC33987 standard; DNA; 959 BP.  
AC 33987;  
AC 33987;  
DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 5039.  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 5039.  
XX Hybridisation assay; genetic mapping; gene expression control;  
KM protein identification; signal transduction pathway;  
KM metabolic pathway; promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX 06-SEP-2000.  
PD 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 14-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.

Alignment Scores:	
Pred. No.:	1.24e-63
Score:	639.00
Percent Similarity:	71.08%
Best Local Similarity:	64.71%
Query Match:	72.12%
Query:	21
Gaps:	3
Length:	955
Matches:	133
Conservative:	13
Mismatches:	25
Indels:	3
Gaps:	4

QY	2	serSermetanpncoglutryAspIryAspIryLeuPheLysLeuLeuLeuIleGlyAspSerGly	21
Db	177	ACAACACGAACTCCTCAATAATGACTATGTGCACACCTTTCGCATGTGGATGTTGTG	236
QY	22	ValGlyLysSerGlyLeuLeuLeuLeuArgPheAlaAspSPTrpYrThrGlnSerTrpIle	41
Db	237	GTTGGAAAGTGGTGCTGCTTCTAAGCTTGGCTGCATGATCTTCACGGTACACTACATC	296
QY	42	serThrIleGlyAlaAspPheLysIleArgTrpIleGluLeuAspGlyLysTrpIleLys	61
Db	297	ACCACCACTGGTGTCGACTTTAAATCGAAGCAGTGCMAACAAAGATGGAAAGACCATCAA	355
QY	62	LeuGlnIle-----	64
Db	357	CTCCACGATTTGGATACGGCAGCGCAGGAACGTTTACAGACATTACTACCACTTACAC	411
QY	65	-----GlnSerPheAsnAsn	69
Db	417	AACAGACCTCATGGAAATTGTGTGACTTATGATGCTACACAGATCTGAAACCTTCAACAC	477
QY	70	ValLysGlnTrpLeuGlnGluIleAspArgTrpAlaSerGluAsnValAsnLysLeuLeu	89
Db	477	GTCACGACATGGCTAAATATAATGACCGCTACGACAGGAGGAAGATGTTAAACAGCTACG	533
QY	90	ValGlyAsnLysCysAspLeuThrTrpLysLysValValAspTrpThrAlaLysGlu	109
Db	537	GTTTGGGAACAAGTGTATATCTCCATCCACAGAAAGTTGTATCCACTGAGACAGCTAAAGCT	594
QY	110	PheAlaAspSerLeuGlyIleLeuPheLeuGlnTrpSerAlaLysAsnAlaThrAsnVal	123
Db	597	TTTCGCTGATGAACTTGGGATCCCATCTTGTGGAAACAAGTGCCTAAGAAATGCTACCAATGTC	656
QY	130	GlnGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThr	145
Db	657	GAGAAAGCTTTCATGGCCCTGACTCTCCGACATCAAAACCCAGAAAGGGG--AACCAACCT	713
QY	150	AlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGly	165

Db 714 GCAGGAGATCTAAGCCACACGCGTCAGATCCGAGACACCCGTTAACGACAA--- 770  
OY 170 GYGlyCysCys 173  
Db 771 TCAGGCTGCTGC 782  
RESULT 11  
AAC35200  
ID AAC35200 standard; DNA: 932 BP.  
XX AAC35200;  
AC AAC35200;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 9347.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135253.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142927.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.



DR WPI, 2001-639362/73.  
DR P-PSDB; ABG19673.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 1; SEQ ID No 19664; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
SQ Sequence 1193 BP; 372 A; 279 C; 213 G; 329 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 8.64e-62 Length: 1193  
Score: 624.00 Matches: 129  
Percent Similarity: 74.57% Conservative: 0  
Best Local Similarity: 74.57% Mismatches: 0  
Query Match: 70.43% Indels: 44  
DB: 23 Gaps: 1  
  
US-09-820-003a-2 (1-173) x AAS83860 (1-1193)  
QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
DB 1003 ATGCCGACGATGCAATCCGAAATGATTATTATTCAGTACTTCGATTGGCGACTCA 944  
QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspSphThrTyrThrGluSerTyr 40  
DB 943 GGGGTTGGAAAGCTTGCTCTCTCTAGGTTGACAGATACATATACAGAAAGCTAC 884  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 883 ATCAGCACAATTGGTGGATTTCATAAATAGAACTATAGATTAGACGGGAAACATC 824  
QY 61 LysLeuGlnIleGluSerPheAsnAsnValLysGlnIlePheGlnGlnIleAspArgTyr 80  
DB 823 AAGCTTCAATA----- 812  
QY 81 AlaSerGluAsnValAlaAsnLysLeuLeuValGlyAsnLysCysAspLeuThrThrLysLys 100  
DB 812 ----- 812  
QY 101 ValValAspTyrThrThrAlaLysGluPheAlaAspSerLeuGlyIleProPheLeuGlu 120  
DB 811 -----GAATTTCGCGATTCCCTTGGCAATTCGCTTTTGGAA 776  
QY 121 ThrSerAlaLysAsnAlaThrAsnValGluGlnSerPheMetThrMetAlaAlaGluIle 140  
DB 775 ACCAGCTCTAAGATGCAACGATGTAAGACAGCTTTCATGACGATGGACGCTGACATT 716  
QY \* 141 LysLysArgMetGlyProGlyAlaThrAlaGlyGlyValGluLysSerAsnValLysIle 160

DB 715 AAAAGCGAATGGGTCCCGGAGACACAGCTGTGGTCTGAGAGCCAAATGTTAAATT 656  
QY 161 GlnSerThrProValLysGlnSerGlyGlyCysCys 173  
DB 655 CAGAGCACTCCAGTCAAGCAGTCAGTGGAGGTTGCTGC 617  
  
RESULT 13  
AAC42684  
ID AAC42684 standard; DNA: 777 BP.  
XX  
XX AAC42684:  
AC  
XX 17-OCT-2000 (first entry)  
DT  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 36471.  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX  
XX EPI033405-A2.  
PN  
XX 06-SEP-2000.  
PD  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 990S-0121825.  
PR 05-MAR-1999; 990S-0123180.  
PR 09-MAR-1999; 990S-0123548.  
PR 23-MAR-1999; 990S-0125788.  
PR 25-MAR-1999; 990S-0126264.  
PR 28-MAR-1999; 990S-0126785.  
PR 01-APR-1999; 990S-0127462.  
PR 06-APR-1999; 990S-0128234.  
PR 08-APR-1999; 990S-0128714.  
PR 16-APR-1999; 990S-0129845.  
PR 19-APR-1999; 990S-0130077.  
PR 21-APR-1999; 990S-0130449.  
PR 23-APR-1999; 990S-0130510.  
PR 23-APR-1999; 990S-0130891.  
PR 28-APR-1999; 990S-0131449.  
PR 30-APR-1999; 990S-0132048.  
PR 30-APR-1999; 990S-0132407.  
PR 04-MAY-1999; 990S-0132484.  
PR 05-MAY-1999; 990S-0132485.  
PR 06-MAY-1999; 990S-0132486.  
PR 07-MAY-1999; 990S-0132487.  
PR 11-MAY-1999; 990S-0132486.  
PR 14-MAY-1999; 990S-0134218.  
PR 14-MAY-1999; 990S-0134219.  
PR 14-MAY-1999; 990S-0134221.  
PR 14-MAY-1999; 990S-0134370.  
PR 18-MAY-1999; 990S-0134768.  
PR 19-MAY-1999; 990S-0134941.  
PR 20-MAY-1999; 990S-0135124.  
PR 21-MAY-1999; 990S-0135353.  
PR 24-MAY-1999; 990S-0135629.  
PR 25-MAY-1999; 990S-0136021.  
PR 27-MAY-1999; 990S-0136392.  
PR 28-MAY-1999; 990S-0136782.  
PR 01-JUN-1999; 990S-0137222.  
PR 03-JUN-1999; 990S-0137528.  
PR 04-JUN-1999; 990S-0137502.  
PR 07-JUN-1999; 990S-0137724.  
PR 08-JUN-1999; 990S-0138094.  
PR 10-JUN-1999; 990S-0138540.  
PR 10-JUN-1999; 990S-0138847.  
PR 14-JUN-1999; 990S-0139119.  
PR 16-JUN-1999; 990S-0139452.





QY	3	serleIsanProgluTyraSPtyleuPheLysLeuLeuilegIlyAspSerGlyVal	22
Db	163	ACCATGAATCTCTGATGACGACTATCTTTCAAGCTCCTGCTTACGGGGATTCGGCGTA	222
QY	23	GlyLysSerCysLeuLeuLeuArgPheAlaAspThrTyrThrGluSerTyrIleSer	42
Db	223	GGCAAGTCTTCTCTCTTTTGGATTCCTGATGATTCCTATGTAGAAGATTACATTAGC	282
QY	43	ThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeu	62
Db	283	ACTATGTGAGTCGATTTTAAATTAAGACGTGGAGAACAGATGGCAAAACAAATTAACTC	342
QY	63	GluIle-----	64
Db	343	CAAAATTGGGACACTGCTGCTCAAGAACGCTTCAGACATTACTAGCACTTACACCT	402
QY	65	-----GluSerPheAsnVal	70
Db	403	GGGGACATGAAATTATATTGTCTACGATGTACAGATGCACAGATGAGAAAGCTCAATTAATGTC	462
QY	71	LysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuVal	90
Db	463	AAGCAATGTTGAGTGAATATGATCGTATGCTAGTACATGTCACATGTCACAAACTCTGTT	522
QY	91	GlyAsnLysCysAspLeuThrThrLysLysValAlaAspTyrThrTrpAlaLysGluPhe	110
Db	523	GGAAACAAGTCTGATCTTACTGAAACAGAACCCATTCCTTATGAAACTGCCAAGCTTT	582
QY	111	AlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaTrpAsnValGlu	130
Db	583	GCCCATGAAACGGGATTCCTTTATGTGAGACTGTGTCACAAAGATGCTACAAACGTGAA	642
QY	131	GlnSerPheMetTrpMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThrAla	150
Db	643	CAGCTTTCATGGCAATGTCTGCATTCCTCAAGAGAAATGCTACCCAAACAGCTGGG	702
QY	151	GlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGlyGly	170
Db	703	AATATGCAAGACACACGACCGCTGACATGAGAGACAGCCTGTGGACAGAAAG--AAC	759
QY	171	GlyCysCys 173	
Db	760	GGCTGCTGC 768	
RESULT 14			
ABN98761			
ID	ABN98761	standard; DNA; 881 BP.	
XX	ABN98761;		
DT	01-AUG-2002	(first entry)	
DE	Arabidopsis thaliana expressed polynucleotide SEQ ID NO 529.		
XX			
KW	Arabidopsis thaliana; plant; insecticide; fungicide; transgenic; stress;		
KW	disease; crop; thale cress; tolerance factor; insect; pathogen;		
KW	nutrition; ds.		
OS	Arabidopsis thaliana.		
XX			
PN	US200203281-A1.		
PD	21-FEB-2002.		
XX			
PF	26-JAN-2001; 2001US-0770445.		
PR	27-JAN-2000; 2000US-178472P.		
XX			
PA	(GORL/) GORLACH J.		
PA	(ANY/) AN Y.		
PA	(HAM/) HAMILTON C M.		
PA	(PRIC/) PRICE J L.		





MON NOV 18 08:42:04 2002

us-09-820-003a-2.rng

Page 19

**OY**    137    ALaAlaGluIleLysrLysrArmeCclProGlyAlaIrmIlaagLygLygluInsyr    156  
         |||       |||||       |||||       |||||       |||||       |||||       |||||  
**Db**     553    ACtGGcAATCAAGACAGAAATGC---AGCAACCTCGAGSAGATCTAAGCACCA    609  
         |||       |||||       |||||       |||||       |||||       |||||       |||||  
**OY**     157    AsnValylLeuInsberThrProAlaLysInserLygLygCysCs    173  
         |||||       |||||       |||||       |||||       |||||       |||||       |||||  
**Db**     610    AGCGTCCAGATCCGAGAACACTTTTAACACAA---TCAGGTCTCTGC    657

Search completed: November 17, 2002, 13:13:16  
Job time : 259 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 17, 2002, 13:08:03; Search time 51 Seconds  
(without alignments)  
1253.443 Million cell updates/sec

Title: US-09-820-003a-2

Perfect score: 886  
Sequence: 1 MSNNPEYDFLKLIGDS.....EKNVKSIOSTPKSGGGCC 173

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 328367 seqs, 184756068 residues

Total number of hits satisfying chosen parameters: 656734

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+.p2n.model -DEV=xih  
-Q=/cgn2.1/USPTO.spool/US09820003/rnucat.13112002.134617.3205/app.query.fasta.1.327  
-DB=Published.Applications.NA -OPWT=fastap -SUFFIX=ripd -MINMATCH=0.1  
-LOOPCL=0 -LOOPTXT=0 -INITTS=bits -START=1 -END=1 -MATRIX=blonsum62  
-TRANS=human40.csl -LIST=45 -DOCLIG=200 -THR.SCORE=pcr -THR.MAX=100  
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USPR=US09820003.ecgn.1.1.20.ecgnat.13112002.134617.3205  
-NCPU=6 -ICPU=3 -NO.XLPHY -NO.MMAP -IARGQUERY -NEG.SCORES=0 -WAIT -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published.Applications.NA.\*

- 1: /cgn2.6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2.6/ptodata/2/pubpna/PTC1\_NEM\_PUB.seq.\*
- 3: /cgn2.6/ptodata/2/pubpna/US06\_NEM\_PUB.seq.\*
- 4: /cgn2.6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2.6/ptodata/2/pubpna/US07\_NEM\_PUB.seq.\*
- 6: /cgn2.6/ptodata/2/pubpna/PTC05\_PUBCOMB.seq.\*
- 7: /cgn2.6/ptodata/2/pubpna/US08\_NEM\_PUB.seq.\*
- 8: /cgn2.6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2.6/ptodata/2/pubpna/US09\_NEM\_PUB.seq.\*
- 10: /cgn2.6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2.6/ptodata/2/pubpna/US10\_NEM\_PUB.seq.\*
- 12: /cgn2.6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2.6/ptodata/2/pubpna/US60\_NEM\_PUB.seq.\*
- 14: /cgn2.6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	886	100.0	1405	US-09-820-003a-1	Sequence 1, Appl1
2	768.5	86.7	925	US-09-967-736-4	Sequence 4, Appl1
3	642	72.5	609	US-09-938-842a-832	Sequence 832, App
4	619.5	69.9	881	US-09-770-445-529	Sequence 529, App

5	487.5	55.0	46050	10	US-09-820-003a-3	Sequence 3, Appl1
6	481	54.3	601	10	US-09-820-003a-36	Sequence 36, Appl1
7	428.5	48.4	651	9	US-09-938-842a-836	Sequence 836, Appl1
8	414	46.7	624	10	US-09-794-257-9	Sequence 9, Appl1
9	414	46.7	1161	10	US-09-794-257-7	Sequence 7, Appl1
10	414	46.7	1161	10	US-09-834-975-879	Sequence 879, Appl1
11	414	46.7	2497	10	US-09-834-975-885	Sequence 885, Appl1
12	414	46.7	2497	10	US-09-834-975-894	Sequence 894, Appl1
13	414	46.7	2497	10	US-09-834-975-896	Sequence 896, Appl1
14	403.5	45.5	1537	10	US-09-925-300-631	Sequence 631, Appl1
15	390.5	44.1	1274	10	US-09-925-302-91	Sequence 91, Appl1
16	344	38.8	538	10	US-09-924-035a-794	Sequence 794, Appl1
17	343	38.7	896	10	US-09-770-445-478	Sequence 478, Appl1
18	333	37.6	639	10	US-09-350-874-66	Sequence 66, Appl1
19	332	37.5	487	10	US-09-864-761-1742	Sequence 1742, Appl1
20	327.5	37.0	1479	10	US-09-822-830a-365	Sequence 365, Appl1
21	327.5	37.0	3257	10	US-09-817-198a-1	Sequence 1, Appl1
22	326.5	36.9	585	10	US-09-917-800a-1461	Sequence 1461, Appl1
23	324	36.6	427	10	US-09-960-352-8911	Sequence 8911, Appl1
24	318	35.9	374	10	US-09-878-574-3637	Sequence 3637, Appl1
25	318	35.9	654	9	US-09-938-842a-2113	Sequence 2113, Appl1
26	316.5	35.7	894	10	US-09-770-445-487	Sequence 487, Appl1
27	316	35.7	3936	10	US-09-919-172-49	Sequence 49, Appl1
28	314.5	35.5	875	12	US-10-051-986-10	Sequence 10, Appl1
29	314.5	35.5	1116	10	US-09-794-257-13	Sequence 13, Appl1
30	314.5	35.5	2674	10	US-09-817-199a-1	Sequence 1, Appl1
31	305.5	34.5	844	10	US-09-770-445-646	Sequence 646, Appl1
32	301.5	34.0	645	9	US-09-938-842a-74	Sequence 74, Appl1
33	300	33.9	654	9	US-09-938-842a-78	Sequence 78, Appl1
34	293	33.1	381	10	US-09-878-574-1365	Sequence 1365, Appl1
35	292.5	33.0	576	10	US-09-794-257-15	Sequence 15, Appl1
36	292.5	33.0	3124	10	US-09-925-302-340	Sequence 340, Appl1
37	292	32.8	179	10	US-09-864-761-18497	Sequence 18497, Appl1
38	291	32.8	273	10	US-09-878-574-9765	Sequence 9765, Appl1
39	285	32.2	348	10	US-09-878-574-2409	Sequence 2409, Appl1
40	285	32.2	771	10	US-09-728-445-652	Sequence 652, Appl1
41	283.5	32.0	1007	10	US-09-822-849a-553	Sequence 553, Appl1
42	281	31.7	405	10	US-09-878-574-3799	Sequence 3799, Appl1
43	280	31.6	645	9	US-09-938-842a-774	Sequence 774, Appl1
44	280	31.6	900	10	US-09-770-445-469	Sequence 469, Appl1
45	276	31.2	676	10	US-09-770-149-374	Sequence 374, Appl1

## ALIGNMENTS

RESULT 1  
US-09-820-003a-1  
Sequence 1, Application US/09820003a  
Patent No. US20020142382A1  
GENERAL INFORMATION:  
APPLICANT: MERRIOV, Genady et al.  
TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
PROTEINS, AND USES THEREOF  
FILE REFERENCE: C1001196  
CURRENT APPLICATION NUMBER: US/09/820,003A  
CURRENT FILING DATE: 2001-03-29  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1405  
TYPE: DNA  
ORGANISM: Homo sapien  
US-09-820-003a-1

Alignment Scores:  
Pred. No.: 9.06e-107  
Score: 886.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 10  
Length: 1405  
Matches: 173  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-820-003A-2 (1-173) x US-09-820-003A-1 (1-1405)

OY	1	MeterSerMetAanProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer	20
Dd	180	ATGTCCAGCAGCATGCCGAATAATGATTATTATTCAAAGTTACTCTTGATGGCGACTCA	239
OY	21	GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrTrpGluSerTyr	40
Dd	240	GGGGTTGGAAAAGTCCTTCCTCTCTTAGGTTGCACATATACATATACAGAAAGCTCAC	289
OY	41	IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle	60
Dd	300	ATCAGCACAAATGGGTGTGGATTAAAAATAAGAACCTATAGAGTTAGCGGGAAAAACATC	359
OY	61	LysLeuGlnIleGluSerPheAsnAsnValLysGlnTrpLeuGlnIuileAspArgTyr	80
Dd	360	AAGCTCAAATAGAGTCCCTTCATTAATGTTAAACAGTGGCGTCGAGAAATAGATCGTTAT	419
OY	81	AlaSerGluAsnValAsnLysLeuLeuValGlyAsnLysCysAspLeuThrThrLysLys	100
Dd	420	GCCAGTGAATAATGTCAACAATTTGTGTGGGAAACAATGTGATGTCGACCACCAAAAGAAA	479
OY	101	ValValAspTyrThrThrAlaLysGluPheAlaAspSerLeuGlyIleProPheLeuGlu	120
Dd	480	GATAGTACACTACACAACAGGGAAGAAATTTGCTGATTCCTTGGAATTCGGTTTTTGAA	539
OY	121	ThreSerAlaLysAsnAlaThrAsnValGluGlnSerThemetThrMetAlaIaGluIle	140
Dd	540	ACCAGTGTAAAGAAATGCAACAAGTGTGAACAGCTTTTCATGACAGATGGAGCGCGAATTT	599
OY	141	LysLysArgMetGlyProGlyValaThrAlaGlyValaGluLysSerAsnValLysIle	160
Dd	600	AAAAAGGAATGGGTCCCGGAGCAACAGCTGTGTCTGTAGAAAGTCCATGTTTAAATTT	659
OY	161	GlnSerThrProValLysGlnSerGlyGlyLysCys	173
Dd	660	CAGAGCACTCCAGTCAAGCAATCCAGGTGGAGGTGTGTCG	698

## RESULT 2

US-09-967-736-4  
; Sequence 4, Application US/09967736  
; Patent No. US20020103340A1

```

; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.

```

Corley, Neil C.  
Shah, Purvi  
TITLE OF INVENTION: RAB PROTEINS  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:

```

;
;      COMPUTER READABLE FORM:
;
;      MEDIUM TYPE: Diskette
;
;      COMPUTER: IBM Compatible
;
;      OPERATING SYSTEM: DOS
;
;      SOFTWARE: FastSeq for Windows Version 2.0
;

```

? FILING DATE: <Unknown>  
 ? ATTORNEY/AGENT INFORMATION:  
 ? NAME: Billings, Lucy J.  
 ? REGISTRATION NUMBER: 36,749  
 ? REFERENCE/DOCKET NUMBER: PR-0367 US  
 ? TELECOMMUNICATION INFORMATION:  
 ? TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

```

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 925 base pair
; TYPE: nucleic acid
; STRADEDNESS: single
; TOPOLOGY: linear
;

```

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

**Alignment Scores:**

Pred. No.:	1.25e+01	length:	92
Score:	768.50	Matches:	15
Percent Similarity:	80.39%	Conservative:	8
Best Local Similarity:	76.47%	Mismatches:	7
Query Match:	86.74%	Indels:	3
DB:	10	Gaps:	2

US-09-820-003A-2 (1-173) x US-09-967-736-4 (1-925)

```

QY      2   SerseletanProglutinyAspTyrLeuPheLysLeuLeuIleGlyAspSergly 21
Db      60   GCGCCCATGAACCCCAGATATGACTACTCTGTTAAGCTGCTTTTGATTGGCGACTAGGC 119
QY      22   ValGlyLysSerCysLeuLeuLeuArgPheAlaAspThrTyrrhgInuSeryTyle 41
Db     120   GTGGCAAGTCATCGTCTGCTCTCGGCTTGCTGATGACACGATCACAGAGAGCTACATC 179
QY      42   SerthrIleGlyValAspPheLysIleArgghrIleIeuLeuAspGlyLysThrIleLys 61
Db     180   AGCACCATCGGGGGTGGAGCTTAAATCGAACCATCGAGCTGATGGCAAACATATAAA 239
QY      62   LeuGlnIle----- 64
Db     240   CTTGAGATCTGGAGACACAGCGGCCAG9AACGCTTCGGACATCACTTCAGCTACTAC 289
QY      65   -----GlutPheAsnAsn 69
Db     300   CGGGGGGCTCATGGCATCATCGTGTGTATACGTCACTGACCGAGATCTCTAGCCAC 359
QY      70   ValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 89
Db     360   GTGAAGCAGTGGCTGCGAAGATTTGACCGCTATCTCCAGCGAAGACGTCAATTAACTCTG 419
QY      90   ValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGlu 109
Db     420   GTGGGCACAAGACGACGACTCACACCAAGAAGSTGGTGCAACACCAAGCCAAAGAG 479
QY     110   PheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnVal 129
Db     480   TTTTCGACGCTCTCGGGGATCCCCCTTTGTGGAGCGAGCGCCAAAGATGCCACCAATGTC 539
QY     130   GluInserPheMetThrMetAlaAlaGluIleLysLysArmeGlyProGlyAlaThr 149
Db     540   GAGCAGCGCTTCATGACCATGCTGCTGTAAATTCAAAAAGCGAGTGGGCGCTGGAGCAGCC 599
QY     150   AlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSergly 169
Db     600   TCTGGGGGCG---GAGCGGCGCCAATCTCAAGATCGACGACACCCCTGTAAAGCGGGCTGCG 656
QY     170   GlyGlyCysCys 173
Db     657   GGTCGCTGTTGC 668

RESULT 3
US-09-938-842A-832
; Sequence 832, Application US/09938842A
; Patent No. US2020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
```

### RESULT 3

```

; Sequence 833, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff

```



APPLICANT : Kieps, Joel  
 APPLICANT : Wang, Xun  
 TITLE OF INVENTION : STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
 FILE REFERENCE : SAME. AND METHODS OF USE  
 CURRENT APPLICATION NUMBER : US/09/938, 842A  
 PRIOR FILING DATE : 2001-08-24  
 PRIOR APPLICATION NUMBER : US 60/227, 866  
 PRIOR FILING DATE : 2000-08-24  
 PRIOR APPLICATION NUMBER : US 60/264, 647  
 PRIOR FILING DATE : 2001-01-16  
 PRIOR APPLICATION NUMBER : US 60/300, 111  
 NUMBER OF SEQ ID NOS : 5379  
 SEQ ID NO 832  
 LENGTH : 609  
 TYPE : DNA  
 ORGANISM : Arabidopsis thaliana  
 US-09-938-842A-832

**Alignment Scores:**  
 Pred. No.: 2,61e-75 Length: 609  
 Score: 642.00 Matches: 133  
 Percent Similarity: 71.29% Conservative: 133  
 Best Local Similarity: 65.84% Mismatches: 24  
 Query Match: 72.46% Indels: 34  
 Gaps: 3

DB: 9 Gaps: 3

US-09-820-003A-2 (1-173) x US-09-938-842A-832 (1-609)  
 QY 4 MetanproctulnGluLeuAspArgTyrAlaSerGluAsnValAsnLysLeuLeuValGly 23  
 Db 1 ATGATCTCTTAATGATATTATTTGTCAAGCTTGTCATGGTGGATCGGTGGGA 60  
 QY 24 LysSerCysLeuLeuLeuAlaAspAspThrTyrThrGluSerTyrIleSerThr 43  
 Db 61 AAGTCGCTGCTGCTTAAGGTTGCTGATGATCCACGTCGATGCATCAACGCC 120  
 QY 44 Tlleglyalaspheylsileargthrillegluleuaspolyleythrileyleu 63  
 Db 121 ATGGGTCTCATCTTAATAATCCGACACGTCGACMACAGATGAAACCATCAACTCAG 180  
 QY 64 Tle----- 64  
 Db 181 ATTGGGATACGCGACGCGAACGTTTCAGAACATTACTACGACTTACTACAGAGA 240  
 QY 65 -----Gluserrpheasnasenvallys 71  
 Db 241 GCCTATGATATTATGCTATGATGATGATACAGATCTGAAAAGCTTCAACAACGCGAAG 300  
 QY 72 Glntfrplengluglileaspargytyralasergluasnvalasnlylsleuleuvalgly 91  
 Db 301 CAATGGCTAAAGAATGACGCCCTATGCAAGTAGAATGTTAAACAAGCTACGTGGG 360  
 QY 92 Asnlyscysaspheuthtthrlyslsyalvalaspyrr-thrThrlalalsgilupha 111  
 Db 361 AACAAAGTGTATCTCACATCCAGAAAGTGTATCCACGACGACGACGACGCTTCCCT 420  
 QY 112 Asperleuglytlaprophelenglhphserlalysasnalatrtanvalglucn 131  
 Db 421 CATGAACCTGGATCCCTTCTTGGAACACAGTCTTAAGTACGACATGCGAAGAA 480  
 QY 132 SerPheletmetelalagluileylslysarngemglyProglyalathraclay 151  
 Db 481 GTTTCATGCGCATCTACTGCTGCAATCAAGACAGCATGAGCC--AGCCAACCTGGAGA 537  
 QY 152 Glyalaglulysersasnvalylsllleglnserthrprovalylsglnserglylely 171  
 Db 538 GGATCTTAAGCCACCAAGCGTCCAGATCCGAGGACACCACTGTTAACCGA---TCAGGC 594  
 QY 172 CysCys 173  
 |||||

---

DB 595 TGCTGC 600

RESULT 4  
 US-09-770-445-529  
 Sequence 529 Application US/09770445  
 Patent No US20020023281A1  
 GENERAL INFORMATION:  
 APPLICANT: Gottsch, Jörn  
 APPLICANT: An, Yong-Qiang  
 APPLICANT: Hamilton, Carol M.  
 APPLICANT: Price, Jennifer L.  
 APPLICANT: Raines, Tracy M.  
 APPLICANT: Yu, Yang  
 APPLICANT: Rameaka, Joshua G.  
 APPLICANT: Page, Amy  
 APPLICANT: Matthew, Abraham V.  
 APPLICANT: Ledford, Brooke L.  
 APPLICANT: Moessner, Jeffrey P.  
 APPLICANT: Haas, William David  
 APPLICANT: Garcia, Carlos A.  
 APPLICANT: Kricker, Meja  
 APPLICANT: Slader, Ted  
 APPLICANT: Davis, Keith R.  
 APPLICANT: Allen, Keith  
 APPLICANT: Hoffman, Neil  
 APPLICANT: Hunban, Patrick  
 TITLE OF INVENTION: Expressed Sequences of Arabidopsis  
 FILE REFERENCE: 2023US (APPA-012PRV)  
 CURRENT APPLICATION NUMBER: US/09/770,445  
 PRIOR FILING DATE: 2001-01-26  
 PRIOR APPLICATION NUMBER: US 60/178,472  
 PRIOR FILING DATE: 2000-01-27  
 NUMBER OF SEQ ID NOS: 999  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 529  
 LENGTH: 881  
 TYPE: DNA  
 ORGANISM: Arabidopsis thaliana  
 US-09-770-445-529

**Alignment Scores:**  
 Pred. No.: 3.94e-72 Length: 881  
 Score: 619.50 Matches: 125  
 Percent Similarity: 69.97% Conservative: 15  
 Best Local Similarity: 61.58% Mismatches: 30  
 Query Match: 69.92% Indels: 33  
 DB: 10 Gaps: 2

US-09-820-003A-2 (1-173) x US-09-770-445-529 (1-881)  
 QY 3 SerMetanproctulnGluLeuAspArgTyrAlaSerGluAsnValAsnLysLeuLeuValGly 22  
 Db 57 ACCATGATCTCTAGACGATATCTTCAAGCTCTGCTTATGGGAGTCTGGCGTA 116  
 QY 23 GlykSerCysLeuLeuLeuAlaAspAspThrTyrThrGluSerTyrIleSer 42  
 Db 117 GCCAGCTGCTGCTTCTTTANATCTCTGATGATCTTCTTAAGTAAAGTACATTAAC 176  
 QY 43 Thrilleglyalaspheylsileargthrillegluleuaspolyleythrileyleu 62  
 Db 177 ACTATGGAGTGATTTTAATAATGAGCTGTGGACAGATGGCAAAACATTAAAGTC 236  
 QY 63 Glnlle----- 64  
 Db 237 CAATTTGGAGACCTGCTGTCAGAACGTTTCAGACGATTAATGACGTTACTACCGT 296  
 QY 65 -----Gluserrpheasnasenval 70  
 Db 297 GGGGACATGGAATTAATTAATGCTACGATGTCACAGATGAGAAAGCTTCAATTAATGTC 356  
 QY 71 LysGlntrplengluglileaspargytyralasergluasnvalasnlylsleuleuval 90  
 |||||



[illegible]

```
Db 61 AAGACCTGCTCCGTGTCCTCCGCTTCTCAGAGAGCGCTTCAACACCACCTTCATCTCCACC 120
Qy 44 IleglValaspPheLysIleArgThrIleglueuaspGlyLysThrIleLysLeuGln 63
Db 121 ATGGGATGATTTTAAATAGACATGAACTAGATGGAAGAAATTAAGCTTCAG 180
Qy 64 Ile----- 64
Db 181 ATATGGGACACAGCGGGTCAAGAAAGATTCGGAACATCACAGACGCTACTACAGAGA 240
Qy 65 -----gluSerPheasnAsnValLys 71
Db 241 GCCATGGCGCATTTATGCTGCTGTATGACATCAAAATGAAAAATCCTTTGACAAATTTAA 300
Qy 72 GlnTrpLeuGlnIleAspArgTyrAlaSerGluasnValAsnLysLeuValGly 91
Db 301 AATTGGATCAGAAACATTGGAAGCAGCCCTCTCCGATGCGAAGAAATGATCTCGGT 360
Qy 92 AsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGluPheAla 111
Db 361 AACAAATGTGATATGATGACAAAGAGCAAGTGCAGAAAGAGGAGGAGGAGGAGCTAGCA 420
Qy 112 AspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln 131
Db 421 ATTGACTATGGGATTAATTCTTGAGACAAGCGCAAAATCAGTGCAGAAATGTAAGAG 480
Qy 132 SerPheMetThrMetAlaIleGluIle-----LysLysArgMetGlyProGly 147
Db 481 GCATTTTTCACCTTGACGACGAGATATATGACAAACATCAACAGAAAAATGAATGACAGC 540
Qy 148 AlaThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGln 167
Db 541 AATTACACGAGGAGCAGGT-----GACACAGTGAAATTAACAGAAAA-CCGATCAAGAA 593
```

## RESULT 9

```
US-09-794-257-7
; Sequence 7, Application US/09794257
; Patent No. US2002009804A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 32705, 23224, 27423, 32700, 32712, NO. US2002009804A1el
; FILE REFERENCE: 35800/209285
; CURRENT APPLICATION NUMBER: US/09/794,257
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/185,606
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)...(641)
US-09-794-257-7
```

## Alignment Scores:

```
Pred. No.: 5.03e-45 Length: 1161
Score: 414.00 Matches: 87
Percent Similarity: 60.508 Conservative: 34
Best Local Similarity: 43.508 Mismatches: 41
Query Match: 46.738 Indels: 39
Db: 10 Gaps: 3
```

US-09-820-003a-2 (1-173) x US-09-794-257-7 (1-1161)

```
Qy 4 MetAsnProGluTyrAspTyrLeuPheLysLeuLeuLeuIleglYasPserGlyValGly 23
Db 18 ATGGCAGAGACGATGATATCTTCAAGCTCTGCTGATCGGCGAGCTGGGGGATGAGC 77
Qy 24 LysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyrIleSerThr 43
```

```
Db 78 AAGACCTGCTCCGTGTCCTCCGCTTCTCAGAGAGCGCTTCAACACCACCTTCATCTCCACC 137
Qy 44 IleglValaspPheLysIleArgThrIleglueuaspGlyLysThrIleLysLeuGln 63
Db 138 ATCGGAATGATTTTAAATAGACATGAACTAGATGGAAGAAATTAAGCTTCAG 197
Qy 64 Ile----- 64
Db 198 ATATGGGACACAGCGGGTCAAGAAAGATTCGGAACATCACAGACGCTACTACAGAGA 257
Qy 65 -----gluSerPheasnAsnValLys 71
Db 258 GCCATGGCGCATTTATGCTGCTGTATGACATCAAAATGAAAAATCCTTTGACAAATTTAA 317
Qy 72 GlnTrpLeuGlnIleAspArgTyrAlaSerGluasnValAsnLysLeuValGly 91
Db 318 AATTGGATCAGAAACATTGGAAGCAGCCCTCTCCGATGCGAAGAAATGATCTCGGT 377
Qy 92 AsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGluPheAla 111
Db 378 AACAAATGTGATATGATGACAAAGAGCAAGTGCAGAAAGAGGAGGAGGAGGAGCTAGCA 437
Qy 112 AspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln 131
Db 438 ATTGACTATGGGATTAATTCTTGAGACAAGCGCAAAATCAGTGCAGAAATGTAAGAG 497
Qy 132 SerPheMetThrMetAlaIleGluIle-----LysLysArgMetGlyProGly 147
Db 498 GCATTTTTCACCTTGACGACGAGATATATGACAAACATCAACAGAAAAATGAATGACAGC 557
Qy 148 AlaThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGln 167
Db 558 AATTACACGAGGAGCAGGT-----GACACAGTGAAATTAACAGAAAA-CCGATCAAGAA 610
```

## RESULT 10

```
US-09-834-975-879
; Sequence 879, Application US/09834975
; Patent No. US2002010815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Hufel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MFI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 879
; LENGTH: 2497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2497)
; OTHER INFORMATION: n = A,T,C or G
US-09-834-975-879
```

## Alignment Scores:

```
Pred. No.: 1.51e-44 Length: 2497
Score: 414.00 Matches: 87
Percent Similarity: 60.508 Conservative: 34
Best Local Similarity: 43.508 Mismatches: 41
Query Match: 46.738 Indels: 39
Db: 10 Gaps: 3
```

US-09-820-003a-2 (1-173) x US-09-834-975-879 (1-2497)

QY	4	MeAspProgluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGlyValGly	23
Db	92	ATGCGAAGACGTCGATTCATCTTCAAGCTCTGCTGATGGGACATCGGGGATAGC	151
QY	24	LysSerScyLeuLeuLeuArgPheAlaSpasPhrTyrThrGluSerTyrIleSerThr	43
Db	152	AAAGACTGCTCTCTGTCGCTTCACAGGAGCCCTTCACAGCACCTTCATCTCCAC	211
QY	44	IleGlyValAspPheLysIleLeuThrIleLeuLeuSclyLysThrIleLysLeuGln	63
Db	212	ATCGAATGTGATTTTAAATAATTAGAACGATACACTAGATCGAATAAAATTAAGCTTCAG	271
QY	64	Ile-----	64
Db	272	ATATGGACACAGCGGTCACGAAAGATTCGGAACATACACACAGCTACTACAGAGA	331
QY	65	-----GluSerPheAsnAsnValLys	71
Db	332	GGCATGGCATTTACGTGCTCATACATCATCAAAATGAAATGCTTTGACATATTAA	391
QY	72	GlnTPLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuValGly	91
Db	392	AATGGATCAGAAACATGTGAAGAGCATCCCTCCGATGCTGAAAGAAATGATCTGGGT	451
QY	92	AsnLysScyAspLeuThrThrLysLysValValAspTyrThrThrAlaLysGluPheAla	111
Db	452	AACAAATGTGATTAAGTAATGACAAAAGACAACTGTCAAAAGAAAGGGGGAAGCTACG	511
QY	112	AspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln	131
Db	512	ATTGACTATGGATTTAAATTTCTTGAGACAAAGCCGCAAAATCCAGGCAAAATGTAGAAAG	571
QY	132	SerPheMetThrMetAlaAlaGluIle-----LysLysArgMetGlyProGly	147
Db	572	GCAATTTTATACATCTGCAGGAGATTAATGTGACAAACCTCACGCAAAAGAAATGACAC	631
QY	148	AlaThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGln	167
Db	632	AATTCAGACGAGACGAGGT-----GACACACTGAAAAATTAACAGAAAA-CGATCAAGAA	684
RESULT	11		
	US-09-834-975-885		
	: Sequence 885, Application US/09834975		
	: Patent No. US20020110815A1		
	: GENERAL INFORMATION:		
	: APPLICANT: Lillie, James		
	: APPLICANT: Brown, Jeffrey		
	: APPLICANT: Bolt, Andrew		
	: APPLICANT: Van Huifel, Christophe		
	: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS		
	: TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY		
	: FILE REFERENCE: MRI-016B		
	: CURRENT APPLICATION NUMBER: US/09/834,975		
	: PRIOR FILING DATE: 2001-04-13		
	: PRIOR APPLICATION NUMBER: 60/197,558		
	: NUMBER OF SEQ ID NOS: 1046		
	: SOFTWARE: FastSeq for Windows Version 4.0		
	: SEQ ID NO: 885		
	: LENGTH: 2497		
	: TYPE: DNA		
	: ORGANISM: Homo sapiens		
	: FEATURE:		
	: NAME/KEY: misc_feature		
	: LOCATION: (1)...(2497)		
	: OTHER INFORMATION: n = A,T,C or G		
	US-09-834-975-885		
Alignment Scores:	1.51e-44	Length: 2497	
Pred. No.:	414.00	Matches: 87	
Score:			

Percent Similarity:	60.508	Conservative:	34
Best Local Similarity:	43.508	Mismatches:	41
Query Match:	46.73%	Indels:	39
DB:	10	Gaps:	3
US-09-820-003A-2 (1-173) x US-09-834-975-885 (1-2497)			
QY 4 MetlanpROGLuTYrAspTYrleuMphelysleuLeuileGLyAspserGLyValGly 23			
DB 92 ATGGCGAAGACGTAGCATATCTCTTCACAGCTCTCTGATCGACATCGGGGATAGC 151			
QY 24 LysSerCysLeuLeuLeuArGpheAlaAspSPThrTYrThlGluSerTYrIleSerThr 43			
DB 152 AAGACGCGCTCTCTGTCGCTTCACAGGACGCTTCACACACCTCTCATCTCCAC 211			
QY 44 lIleGlyAlaSPheLysIleArGThrIleGLuLeuAspGlyLySThrIleLysLeuGln 63			
DB 212 ATCGGAATTTGATTTTAAATTTAGACATGATGACTGATGGAAAGAAATTTAACTTCAG 271			
QY 64 lIle----- 64			
DB 272 ATATGGGACACACGCGGTGAGAAAGATTCCGAAATACGACAGCGTACTACAGGA 331			
QY 65 -----GluserPheAsnValIys 71			
DB 332 GCCATGGGCGATTATGCTGCTATGACATCCAAATGAAATAATCTTGTACATATTAA 391			
QY 72 GlnTrpLeuGluIleAspArGTYrIleSerGLyAsnValAsnLysleuLeuValGly 91			
DB 392 AATTGGATCGAAGAACTTGAAGAGCATGCTCTCGCATGTCGAAAGAAATGATCTGTGG 451			
QY 92 AsnLysCysAspLeuThrThrlLysValValAlaSPYrThrThrlAlaGluPheAla 111			
DB 452 AACCAATGTGATATGATGACAAAAGAAAGAACTCTCAAAGAAAGAGGAGAGACTAGCA 511			
QY 112 AsperIeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln 131			
DB 512 ATTGACATGGGATTAATCTTGGAGACAGCGCAAAATCCAGTCAATGTGTGAAGAG 571			
QY 132 SerPheMetThrMetAlaAlaGluIle-----LysLysArgMetGlyProGly 147			
DB 572 GCATTTTTTTCACCTTGACAGAGATTAATGTGACAAACTCAACAGAAATGATGACAC 631			
QY 148 AlaThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGln 167			
DB 632 AATTCACAGCAGGACGAGT-----GAGCCAGTGAATAATMACAGAAA-CCGATCAAGAA 684			
RESULT 12			
US-09-834-975-894			
: Sequence 894, Application US/09834975			
: Patent No. US20020110815A1			
: GENERAL INFORMATION:			
: APPLICANT: Lillie, James			
: APPLICANT: Brown, Jeffrey			
: APPLICANT: Bolt, Andrew			
: APPLICANT: Van Hufel, Christophe			
: TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS			
: TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY			
: FILE REFERENCE: MFI-0168			
: CURRENT APPLICATION NUMBER: US/09/834,975			
: PRIOR FILING DATE: 2001-04-13			
: PRIOR APPLICATION NUMBER: 60/197,538			
: NUMBER OF SEQ ID NOS: 1046			
: SOFTWARE: FASTSED for Windows Version 4.0			
: SEQ ID NO 894			
: LENGTH: 2497			
: TYPE: DNA			
: ORGANISM: Homo sapiens			
: FEATURE:			
: NAME/KEY: misc_feature			
: LOCATION: (1)...(2497)			

; OTHER INFORMATION: n = A,T,C or G  
US-09-834-975-894

## Alignment Scores:

Pred. No.:	1.51e-44	Length:	2497
Score:	414.00	Matches:	87
Percent Similarity:	60.50%	Conservative:	34
Best Local Similarity:	43.50%	Mismatches:	41
Query Match:	46.73%	Indels:	39
DB:	10	Gaps:	3

US-09-820-003a-2 (1-173) x US-09-834-975-894 (1-2497)

```
QY 4 MetasnProgiuTyrrAspTyrlEuPhelYsleuLeuilegIlyAspserGlyValGly 23
|||
DB 92 ATGCGAAGAGCGATGATATCTTCAAGCTCCTGCTGATCGGCGAGCTCGGGGTAGGC 151
QY 24 LysSerCysleuLeuValgPhelaAspAspThrTyrrThrGuserrTyrlleSerThr 43
|||
DB 152 AAGACCTGCTCCCTGCTCCGCTTCTCAGAGAGCGCTTCAACACACCTTCATCTCCACC 211
QY 44 lIleGlyValasphelysIleargThrIleGlyleuAspGlyLysThrIleYsleuGln 63
|||
DB 212 ATCGGAATGTATTTAAATTAGACATGACTAGATGGAAGAAATTAACTTCAG 271
QY 64 lIe-----
|||
DB 272 ATATGGGACACAGCGGGTCAGAAAGATTCGAAACATCAGACGCTACTACAGAGA 331
QY 65 -----
|||
DB 332 GCCATGGGCAATTATGCTGGTCTATGACATCAAAATGAAATTCCTTGACAAATTTAA 391
QY 72 GlnTrpLeuGlnIuileAspArgTyrlAlaserGluAsnValasIleuValGly 91
|||
DB 392 AATTGGATCAGAAACATTGGAAGACATGCCCTTCCGATGTGCAAGATATCTCGGT 451
QY 92 AsnLysCysAspleuThrThrLysValValAspTyrrThrAlaYsGluPhela 111
|||
DB 452 AACCAATGTATGATGATGACAAAGACAGTGTCAAAAGAGAGGAGGAGAACTGCA 511
QY 112 AspserLeuGlyIleProphleuGlnThrSerAlaYsAsnAlaThrAsnValGluGln 131
|||
DB 512 ATTGACTATGGGATTAATCTTGAGACAAAGCGCAAAATCCAGTGCAGATGGAAGAG 571
QY 132 SerPheMetThrMetAlaIaGluIle-----LysLysArgMetGlyProGly 147
|||
DB 572 GCATTTTACACTTGCACGAGATATATGCAAAACTCAACAGAAAATTAATGACAGC 631
QY 148 AlaThrAlaGlyIaIaGluLysSerAsnValYsIleGlnSerThrProValYsGln 167
|||
DB 632 AATTCACAGACAGACAGGT-----GACACAGTGAAATTAACAGAAA-CCGATCAAGAA 684
```

## RESULT 13

US-09-834-975-896  
; Sequence 896, Application US/09834975  
; Patent No. US20020110815A1  
; GENERAL INFORMATION:  
; APPLICANT: lillie, James  
; APPLICANT: Brown, Jeffrey  
; APPLICANT: Bolt, Andrew  
; APPLICANT: Van Huifel, Christophe  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS  
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY  
; FILE REFERENCE: MRI-016B  
; CURRENT APPLICATION NUMBER: US/09/834, 975  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/197,538  
; NUMBER OF SEQ ID NOS: 1046  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 896

; LENGTH: 2497  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(2497)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-834-975-896

## Alignment Scores:

Pred. No.:	1.51e-44	Length:	2497
Score:	414.00	Matches:	87
Percent Similarity:	60.50%	Conservative:	34
Best Local Similarity:	43.50%	Mismatches:	41
Query Match:	46.73%	Indels:	39
DB:	10	Gaps:	3

US-09-820-003a-2 (1-173) x US-09-834-975-896 (1-2497)

```
QY 4 MetasnProgiuTyrrAspTyrlEuPhelYsleuLeuilegIlyAspserGlyValGly 23
|||
DB 92 ATGCGAAGAGCGATGATATCTTCAAGCTCCTGCTGATCGGCGAGCTCGGGGTAGGC 151
QY 24 LysSerCysleuLeuValgPhelaAspAspThrTyrrThrGuserrTyrlleSerThr 43
|||
DB 152 AAGACCTGCTCCCTGCTCCGCTTCTCAGAGAGCGCTTCAACACACCTTCATCTCCACC 211
QY 44 lIleGlyValasphelysIleargThrIleGlyleuAspGlyLysThrIleYsleuGln 63
|||
DB 212 ATCGGAATGTATTTAAATTAGACATGACTAGATGGAAGAAATTAACTTCAG 271
QY 64 lIe-----
|||
DB 272 ATATGGGACACAGCGGGTCAGAAAGATTCGAAACATCAGACGCTACTACAGAGA 331
QY 65 -----
|||
DB 332 GCCATGGGCAATTATGCTGGTCTATGACATCAAAATGAAATTCCTTGACAAATTTAA 391
QY 72 GlnTrpLeuGlnIuileAspArgTyrlAlaserGluAsnValasIleuValGly 91
|||
DB 392 AATTGGATCAGAAACATTGGAAGACATGCCCTTCCGATGTGCAAGATATCTCGGT 451
QY 92 AsnLysCysAspleuThrThrLysValValAspTyrrThrAlaYsGluPhela 111
|||
DB 452 AACCAATGTATGATGATGACAAAGACAGTGTCAAAAGAGAGGAGGAGAACTGCA 511
QY 112 AspserLeuGlyIleProphleuGlnThrSerAlaYsAsnAlaThrAsnValGluGln 131
|||
DB 512 ATTGACTATGGGATTAATCTTGAGACAAAGCGCAAAATCCAGTGCAGATGGAAGAG 571
QY 132 SerPheMetThrMetAlaIaGluIle-----LysLysArgMetGlyProGly 147
|||
DB 572 GCATTTTACACTTGCACGAGATATATGCAAAACTCAACAGAAAATTAATGACAGC 631
QY 148 AlaThrAlaGlyIaIaGluLysSerAsnValYsIleGlnSerThrProValYsGln 167
|||
DB 632 AATTCACAGACAGACAGGT-----GACACAGTGAAATTAACAGAAA-CCGATCAAGAA 684
```

## RESULT 14

US-09-925-300-631  
; Sequence 631, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Kuden,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270



Mon Nov 18 08:42:04 2002

us-09-820-003a-2.rnpb

Page 10

Search completed: November 17, 2002, 14:31:59  
Job time : 62 secs

---



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 13:01:57 ; Search time 43 Seconds  
(without alignments)  
386.773 Million cell updates/sec

Title: us-09-820-003a-2

Perfect score: 886  
Sequence: 1 MSSNMPEDYLFKLLIGDS.....EKSNNKIQSTPVKQSGGCC 173

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR73:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	860	97.1	205	1 TVHUTP	GTP-binding protein
2	860	97.1	205	1 TVDSTP	GTP-binding protein
3	860	97.1	205	1 TVKSTP	GTP-binding protein
4	849	95.8	205	1 TVRSTP	GTP-binding protein
5	812.5	91.7	201	2 D38625	GTP-binding protein
6	771.5	87.1	201	2 S06147	GTP-binding protein
7	733	82.7	205	2 S38339	GTP-binding protein
8	671	75.7	202	2 JE0318	GTP-binding protein
9	660	74.5	205	2 T33781	hypothetical protein
10	658.5	74.3	203	2 S30096	GTP-binding protein
11	646.5	73.0	203	2 JC4105	GTP-binding protein
12	646.5	73.0	203	2 T50323	GTP-binding protein
13	642.5	72.5	203	2 JC1247	GTP-binding protein
14	631.5	71.3	201	2 UC5337	GTP-binding protein
15	627.5	70.8	203	2 S34253	GTP-binding protein
16	623.5	70.4	206	2 S04590	GTP-binding protein
17	619.5	69.9	258	2 B86153	GTP-binding protein
18	619	69.9	202	2 S41430	GTP-binding protein
19	618.5	69.8	203	2 B38202	GTP-binding protein
20	613.5	69.1	202	2 S38740	GTP-binding protein
21	612	69.1	221	2 H71444	GTP-binding protein
22	598	67.5	218	2 T07609	GTP-binding protein
23	590.5	66.6	202	2 S72515	GTP-binding protein
24	587	66.3	201	2 S39655	GTP-binding protein
25	574	64.8	206	2 T14391	GTP-binding protein
26	572.5	64.6	196	2 PS0279	GTP-binding protein
27	571	64.4	206	1 TVBQ2	GTP-binding protein
28	523	59.0	208	2 A38202	GTP-binding protein
29	468.5	52.9	208	2 A34716	GTP-binding protein

30	461	52.0	203	2 B34716	GTP-binding protein
31	447	50.5	216	2 T48378	GTP-binding protein
32	442	49.9	215	2 T14565	GTP-binding protein
33	439.5	49.6	215	2 S57478	GTP-binding protein
34	438.5	49.5	216	2 S06640	GTP-binding protein
35	435.5	49.2	200	2 S12790	GTP-binding protein
36	430.5	48.6	215	2 S57462	GTP-binding protein
37	429.5	48.5	215	2 S57471	GTP-binding protein
38	429.5	48.5	222	2 T14405	GTP-binding protein
39	428.5	48.4	216	2 T45901	GTP-binding protein
40	426.5	48.1	215	2 S57474	GTP-binding protein
41	425.5	48.0	216	2 S33900	GTP-binding protein
42	418	47.2	207	2 B49647	GTP-binding protein
43	418	47.2	207	2 B36364	GTP-binding protein
44	410.5	46.3	206	2 T17851	GTP-binding protein
45	409.5	46.2	209	2 B38625	GTP-binding protein

## ALIGNMENTS

## RESULT 1

GTP-binding protein Rabi1 - human  
TVHUTP  
N:Alternate names: protein DKFZp564B163.1  
C:Species: Homo sapiens (man)  
C>Date: 29-Jun-1990 #sequence\_revision 06-Dec-1996 #text\_change 19-Jan-2001  
C:Accession: A34323; T08698  
R:Zahraoui, A.; Touchot, N.; Chardin, P.; Tavallan, A.  
J. Biol. Chem. 264, 12394-12401, 1989  
A:Title: The human Rabi1 gene encodes a family of GTP-binding proteins related to yeast.  
A:Reference number: A34323; MUID:89308668; PMID:2501306  
A:Accession: A34323  
A:Molecule type: mRNA  
A:Residues: 1-205 <ZMW>  
A:Cross-references: GB:04941; GB:M28209; NID:9550059; PIDN:AAA60240.1; PID:9550060  
A:Experimental source: pheochromocytoma  
R:Wambitt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, March 1999  
A:Reference number: Z16471  
A:Accession: T08698  
A:Molecule type: mRNA  
A:Residues: 1-64,141-205 <ZMW>  
A:Cross-references: EMBL:AL050268  
A:Experimental source: fetal brain; clone DKFZp564B163  
C:Genetics:  
A:Gene: GDB:RAB1  
A:Cross-references: GDB:118857; OMIM:179508  
A:Map position: 4P15.31-4P15.31  
A>Note: DKFZp564B163.1  
C:Function:  
A:Description: probably involved in protein transport from the endoplasmic reticulum  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipid protein; membran  
tein  
F:1-205/Product: GTP-binding protein Rabi1 #status predicted <NAT>  
F:1-64,141-205/Product: GTP-binding protein Rabi1, splice variant #status predicted <K  
F:12-127/Domain: translation elongation factor Tu homology <ETU>  
F:18-25/Region: nucleotide-binding motif A (P-loop)  
F:62-67/Region: nucleotide-binding motif B  
F:124-127/Region: GTP-binding NKXD motif  
F:154-156/Region: GTP-binding NKXD motif  
F:124,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #  
F:124/Binding site: phosphate (Ser) (covalent) (by cdc2 kinase) #status predicted  
F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status experimental

Query Match 97.1% Score 860; DB 1; Length 205;  
Best Local Similarity 84.4% Pred. No. 2,2e+62;

Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MSSNMPEDYLFKLLIGDSGVKSCLLRFADDTYESYSTGDEFKRIEIDGKTI 60  
DB 1 MSSNMPEDYLFKLLIGDSGVKSCLLRFADDTYESYSTGDEFKRIEIDGKTI 60

QY 61 KLOI-----ESFNNVKOMLOEIDRYASENVKL 88  
|||  
Db 61 KLOIWDIAGORFRRTITSSYRGAHGIIVYDVTDQSEFNNVKOMLOEIDRYASENVKL 120  
QY 89 LVGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPGA 148  
Db 121 LVGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPGA 180  
QY 149 TAGGAEKSNVNIQSTPVKQSGGCC 173  
Db 181 TAGGAEKSNVNIQSTPVKQSGGCC 205

## RESULT 2

## TVDCYP

Q:Species: Canis lupus familiaris (dog)  
C:Date: 19-Feb-1994 #sequence\_revision 06-Dec-1996 #text\_change 19-Jan-2001  
C:Accession: S19104; A36364; S15600  
R:Zerial, M.  
Submitted to the EMBL Data Library, August 1990  
A:Reference number: S19104  
A:Accession: S19104  
A:Molecule type: mRNA  
A:Residues: 1-205 <ZER>  
A:Cross-references: EMBL:X56384  
R:Chavrier, P.; Vigneron, M.; Sander, C.; Simons, K.; Zerial, M.  
Mol. Cell. Biol. 10, 6578-6585, 1990  
A:Title: Molecular cloning of IPT1/SEC4-related cDNAs from an epithelial cell line.  
A:Reference number: A36364; MUID:91061765; PMID:2123294  
A:Accession: A36364  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-157, 'EK', 160-205 <CHA>  
A:Cross-references: GB:X56384; NID:9913  
C:Function:  
A:Description: probably involved in protein transport from the endoplasmic reticulum thru  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipoprotein; membrane protein

F:12-127/Domain: translation elongation factor Tu homology <ETU>  
F:18-23/Region: nucleotide-binding motif A (P-loop)  
F:62-67/Region: nucleotide-binding motif B  
F:124-127/Region: GTP-binding NKXD motif  
F:154-156/Region: GTP-binding SAK/L motif  
F:24,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
F:194/Binding site: phosphate (Ser) (covalent) (by cd2 kinase) #status predicted  
F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 97.1%; Score 860; DB 1; Length 205;  
Best Local Similarity 84.4%; Pred. No. 2.2e-62;  
Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MSSNNPEYDIFKLLIGDSGVGKSCLLRPADTYTESYSTIGVDFKRTILDGKI 60  
Db 1 MSSNNPEYDIFKLLIGDSGVGKSCLLRPADTYTESYSTIGVDFKRTILDGKI 60  
QY 61 KLOI-----SEFNNVKOMLOEIDRYASENVKL 88  
|||  
Db 61 KLOIWDIAGORFRRTITSSYRGAHGIIVYDVTDQSEFNNVKOMLOEIDRYASENVKL 120  
QY 89 LVGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPGA 148  
Db 121 LVGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPGA 180  
QY 149 TAGGAEKSNVNIQSTPVKQSGGCC 173  
Db 181 TAGGAEKSNVNIQSTPVKQSGGCC 205

## RESULT 3

## TVKSTP

... 9

GTP-binding protein ypt1 - mouse  
N:Alternate names: GTP-binding protein Rab1; ras-related protein ypt1; transforming p  
C:Species: Mus musculus (house mouse)  
C:Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 19-Jan-2001  
C:Accession: S05551; S06285  
R:Wichmann, H.; Disela, C.; Haubruck, H.; Gallwitz, D.  
Nucleic Acids Res. 17, 6737-6738, 1989  
A:Title: Nucleotide sequence of the mouse ypt1 gene encoding a ras-related GTP-bindin  
A:Reference number: S05551; MUID:89386011; PMID:2506528  
A:Accession: S05551

A:Molecule type: DNA

A:Residues: 1-205 <WIC>

A:Cross-references: EMBL:X15744; NID:955458; PIDN:CAA33760.1; PID:9763158  
R:Haubruck, H.; Disela, C.; Wagner, P.; Gallwitz, D.  
EMBO J. 6, 4049-4053, 1987

A:Title: The ras-related ypt protein is an ubiquitous eukaryotic protein: isolation a  
A:Reference number: S06285; MUID:88166649; PMID:3127202

A:Accession: S06285

A:Molecule type: mRNA

A:Residues: 1-205 <HAU>

C:Genetics:

A:Gene: ypt1

A:Introns: 8/2; 32/3; 64/3; 96/3; 140/3

C:Function:  
A:Description: probably involved in protein transport from the endoplasmic reticulum  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology

C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipoprotein; membran  
tein

F:12-127/Domain: translation elongation factor Tu homology <ETU>  
F:18-25/Region: nucleotide-binding motif A (P-loop)

F:62-67/Region: nucleotide-binding motif B  
F:124-127/Region: GTP-binding NKXD motif

F:154-156/Region: GTP-binding SAK/L motif  
F:24,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #

F:194/Binding site: phosphate (Ser) (covalent) (by cd2 kinase) #status predicted  
F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 97.1%; Score 860; DB 1; Length 205;  
Best Local Similarity 84.4%; Pred. No. 2.2e-62;

Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MSSNNPEYDIFKLLIGDSGVGKSCLLRPADTYTESYSTIGVDFKRTILDGKI 60  
Db 1 MSSNNPEYDIFKLLIGDSGVGKSCLLRPADTYTESYSTIGVDFKRTILDGKI 60  
QY 61 KLOI-----SEFNNVKOMLOEIDRYASENVKL 88  
|||  
Db 61 KLOIWDIAGORFRRTITSSYRGAHGIIVYDVTDQSEFNNVKOMLOEIDRYASENVKL 120  
QY 89 LVGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPGA 148  
Db 121 LVGNKCDLTTKKVVDYTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPGA 180  
QY 149 TAGGAEKSNVNIQSTPVKQSGGCC 173  
Db 181 TAGGAEKSNVNIQSTPVKQSGGCC 205

## RESULT 4

## TVRTYP

GTP-binding protein Rab1 - rat  
N:Alternate names: transforming protein ypt1 homology  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 03-Aug-1992 #sequence\_revision 06-Dec-1996 #text\_change 19-Jan-2001  
C:Accession: A39963  
R:Touchot, N.; Chardin, P.; Tavilian, A.  
Proc. Natl. Acad. Sci. U.S.A. 84, 8210-8214, 1987  
A:Title: Four additional members of the ras gene superfamily isolated by an oligonuc  
A:Reference number: A39963; MUID:88068563; PMID:3317403  
A:Accession: A39963  
A:Molecule type: mRNA  
A:Residues: 1-205 <TOD>  
A:Cross-references: GB:J02998; NID:9206552; PIDN:AAA42006.1; PID:9206553

C:Function: A:Description: probably involved in protein transport from the endoplasmic reticulum th C:Species: rat C:Keywords: endoplasmic reticulum; Golgi apparatus; GTP binding; lipoprotein; membrane F

F:12-127/Domain: translation elongation factor Tu homology <ETU>  
F:18-25/Region: nucleotide-binding motif A (P-loop)  
F:62-67/Region: nucleotide-binding motif B  
F:124-127/Region: GTP-binding NKX/L motif  
F:154-156/Region: GTP-binding SAK/L motif  
F:24,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
F:194/Binding site: phosphatase (Ser) (covalent) (by cdc2 kinase) #status predicted  
F:204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 95.8%; Score 849; DB 1; Length 205;  
Best Local Similarity 83.4%; Pred. No. 1,7e-61;  
Matches 171; Conservative 0; Mismatches 2; Indels 32; Gaps 1;

QY 1 MSNMPEYDLFKLLIGDSGKSCLLRFADDTYTESYSTIGVDFKIRTELDGKTI 60  
D 1 MSNMPEYDLFKLLIGDSGKSCLLRFADDTYTESYSTIGVDFKIRTELDGKTI 60  
QY 61 KIQI-----ESFNNVKOMLOEIDRYASBNKLV 88  
D 61 KIQIMDPAQGERFRTTSSYRGAHGIYVDVDTQESFNNVKOMLOEIDRYASBNKLV 120  
QY 89 LVGKKCDLTTRKKVYDTTAKFEADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPG 148  
D 121 LVGKKCDLTTRKKVYDTTAKFEADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPG 180  
QY 149 TAGAEEKSNVKIQTSPVKOSGGGCC 173  
D 181 TAGAEEKSNVKIQTSPVKOSGGGCC 205

RESULT 5  
D38625  
GTP-binding protein o-rab1 - electric ray (Discopyge omata)  
C:Species: Discopyge omata  
C>Date: 23-Aug-1991 #sequence\_rev150 23-Aug-1991 #text\_change 02-Feb-2001  
C:Accession: D38625  
R:Ngase, J.K.; Eitelink, L.A.; Scheffer, R.H.  
J. Biol. Chem. 266, 2675-2680, 1991  
A:Title: A family of ras-like GTP-binding proteins expressed in electrotomotor neurons.  
A:Reference number: A38625; MUID:91115900; PMID:1899244  
A:Accession: D38625  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-201 <NGS>  
A:Cross-references: GB:M38393  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;  
F:9-124/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:121-124/Region: GTP-binding NKX/L motif  
F:150-152/Region: GTP-binding SAK/L motif  
F:21,22,40,121,122,124,150/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
F:200,201/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 91.7%; Score 812.5; DB 2; Length 201;  
Best Local Similarity 81.2%; Pred. No. 1.5e-58;  
Matches 164; Conservative 3; Mismatches 2; Indels 33; Gaps 2;

QY 4 MNPEYDLFKLLIGDSGKSCLLRFADDTYTESYSTIGVDFKIRTELDGKTI 63  
D 1 MNPEYDLFKLLIGDSGKSCLLRFADDTYTESYSTIGVDFKIRTELDGKTI 60  
QY 64 I-----ESFNNVKOMLOEIDRYASBNKLV 91  
D 64 IMDPAQGERFRTTSSYRGAHGIYVDVDTQESFNNVKOMLOEIDRYASBNKLV 120  
QY 121 LVGKKCDLTTRKKVYDTTAKFEADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPG 180  
QY 149 TAGAEEKSNVKIQTSPVKOSGGGCC 173  
D 181 TAGAEEKSNVKIQTSPVKOSGGGCC 205

D 121 NKCDL-TRKKVYDTTAKFEADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPGATSG 179  
QY 152 GAEEKSNVKIQTSPVKOSGGGCC 173  
D 180 GSKEKSNVKIQTSPVKOSGGGCC 201

RESULT 6  
S06147  
GTP-binding protein rab1b - rat  
N:Alternate names: ras-related protein rab1b  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 30-Sep-1991 #sequence\_rev150 30-Sep-1991 #text\_change 02-Feb-2001  
C:Accession: S06147; S03189  
R:Touchot, N.; Zahraoui, A.; Vleth, E.; Tavlihan, A.  
FEBS Lett. 256, 79-84, 1989  
A:Title: Biochemical properties of the YPT-related rab1b protein. Comparison with rab  
A:Reference number: S06147; MUID:90033316; PMID:2509243  
A:Accession: S06147  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-201 <NOU>  
R:Vleth, E.; Touchot, N.; Zahraoui, A.; Tavlihan, A.  
Nucleic Acids Res. 17, 1770, 1989  
A:Title: Nucleotide sequence of a rat cDNA: RAB1B, encoding a RAB1-YPT related protel  
A:Reference number: S03189; MUID:89160341; PMID:2493636  
A:Accession: S03189  
A:Molecule type: mRNA  
A:Residues: 1-95, 'YA', 98-201 <VTE>  
A:Cross-references: EMBL:X13905; NID:957005; PIDN:CA032105.1; PID:957006  
C:Genes: rab1b  
C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop  
F:9-124/Domain: translation elongation factor Tu homology <ETU>  
F:15-22/Region: nucleotide-binding motif A (P-loop)  
F:121-124/Region: GTP-binding NKX/L motif  
F:151-153/Region: GTP-binding SAK/L motif  
F:21,22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #  
F:200,201/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 87.1%; Score 771.5; DB 2; Length 201;  
Best Local Similarity 77.2%; Pred. No. 3e-55;  
Matches 156; Conservative 6; Mismatches 7; Indels 33; Gaps 2;

QY 4 MNPEYDLFKLLIGDSGKSCLLRFADDTYTESYSTIGVDFKIRTELDGKTI 63  
D 1 MNPEYDLFKLLIGDSGKSCLLRFADDTYTESYSTIGVDFKIRTELDGKTI 60  
QY 64 I-----ESFNNVKOMLOEIDRYASBNKLV 91  
D 64 IMDPAQGERFRTTSSYRGAHGIYVDVDTQESFNNVKOMLOEIDRYASBNKLV 120  
QY 121 LVGKKCDLTTRKKVYDTTAKFEADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPGATAG 151  
D 121 NKSDLTTRKKVYDTTAKFEADSLGIPLETSAKNATNVEQSFMTMAEIKRMGPGATAG 180  
QY 152 GAEEKSNVKIQTSPVKOSGGGCC 173  
D 181 G-ERENLKTIDSTPVKASGGCC 201

RESULT 7  
S38339  
GTP-binding protein rab1 - great pond snail  
C:Species: Lymnaea stagnalis (great pond snail)  
C>Date: 13-Jan-1995 #sequence\_rev150 13-Jan-1995 #text\_change 02-Feb-2001  
C:Accession: S38339; S32206  
R:Agteberg, M.; van Die, I.; Yang, H.; Andriessen, J.A.; van Teetering, A.; van den E  
Eur. J. Biochem. 217, 241-246, 1993  
A:Title: Isolation and characterization of three cDNAs coding for Rab proteins from t  
A:Reference number: S38339; MUID:94039042; PMID:8223561  
A:Accession: S38339

A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-205 <AGT>  
A;Cross-references: EMBL:X72688; NID:g288933; PIDN:CA51233.1; PID:g288934  
C;Genetics:  
A;Gene: rab1  
C;Superfamily: ras transforming protein; translation elongation factor Tu homology  
C;Keywords: GTP binding; lipoprotein; nucleotide binding; P-loop; prenylated cysteine  
F;12-127/Domain: translation elongation factor Tu homology <ETU>  
F;18-25/Region: nucleotide-binding motif A (P-loop)  
F;124-127/Region: GTP-binding NKXD motif  
F;154-156/Region: GTP-binding SAK/L motif  
F;24,25,43,124,125,127,134/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
F;204,205/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 82.7%; Score 733; DB 2; Length 205;  
Best Local Similarity 73.8%; Pred. No. 4e-52;  
Matches 152; Conservative 7; Mismatches 13; Indels 34; Gaps 3;

QY 1 MSSMNPEDYLFKLLIGDSGVGKSCLLLRPADDTYESYISTIGVDFKIRTIELDGKTI 60  
DB 1 MSFMNPEDYLFKLLIGDSGVGKSCLLLRPADDTYESYISTIGVDFKIRTIELDGKTI 60  
QY 61 KLOI-----ESFNNKQWLOEIDRYASENNKL 88  
DB 61 KLOIMDTAGGERFTITSSYRGAGIIVYDVTDOESFNNKQWLOEIDRYASENNKL 120  
QY 89 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGP 148  
DB 121 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGP-1 179  
QY 149 TAGGAESKNVKI-QSTPVKQSGGCC 173  
DB 180 TAASDSKPSVKINSSTPVSAKNGGCC 205

RESULT 8  
JE0318  
GTP-binding protein rabd - silkworm  
C;Species: Bombyx mori (silkworm)  
C;Date: 05-Feb-1999 #sequence\_revision 05-Feb-1999 #text\_change 02-Feb-2001  
C;Accession: JE0318  
R;Uno, T.; Ueno, M.; Nakajima, A.; Shirai, Y.; Aizono, Y.  
Biosci. Biotechnol. Biochem. 62, 1885-1891, 1998  
A;Title: Molecular cloning of cDNA for Brab from the brain of Bombyx mori and biochemica  
A;Reference number: JE0318; MUID:99053143; PMID:9836423  
A;Accession: JE0318  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-202 <UNO>  
A;Cross-references: GB:AF013572  
C;Superfamily: ras transforming protein; translation elongation factor Tu homology  
C;Keywords: GTP binding; nucleotide binding; P-loop  
F;9-124/Domain: translation elongation factor Tu homology <ETU>  
F;15-22/Region: nucleotide-binding motif A (P-loop)  
F;121-124/Region: GTP-binding NKXD motif  
F;151-153/Region: GTP-binding SAK/L motif  
F;21,22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

Query Match 75.7%; Score 671; DB 2; Length 202;  
Best Local Similarity 66.5%; Pred. No. 4e-47;  
Matches 135; Conservative 18; Mismatches 16; Indels 34; Gaps 3;

QY 4 MNPEYDLFKLLIGDSGVGKSCLLLRPADDTYESYISTIGVDFKIRTIELDGKTI 63  
DB 1 MNPEYDLFKLLIGDSGVGKSCLLLRPADDTYESYISTIGVDFKIRTIELDGKTI 60  
QY 64 I-----ESFNNKQWLOEIDRYASENNKL 91  
DB 61 IMPTAGGERFTITSSYRGAGIIVYDCTDQDSFNNKQWLEIDRYACDNNKLLV 120  
QY 92 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGP 151  
DB 92 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGP 151

DB 121 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPST-G 179  
QY 152 GAESKNVKI-QSTPVKQSGGCC 173  
DB 180 AAPAGHVKIDGGPIDGKSSCC 202

RESULT 9  
T33781  
hypothetical protein C39F7.4 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 11-Jan-2000  
C;Accession: T33781  
R;Maggi, L.; Scheet, P.; Dubbelde, C.  
submitted to the EMBL Data Library, October 1998  
A;Description: The sequence of C. elegans cosmid C39F7.  
A;Reference number: Z21407  
A;Accession: T33781  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-205 <MAG>  
A;Cross-references: EMBL:AF10310; PIDN:AC69218.1; GSPDB:GN00023; CESP:C39F7.4  
A;Experimental source: strain Bristol N2; clone C39F7  
C;Genetics:  
A;Gene: CESP:C39F7.4  
A;Map position: 5  
A;Introns: 8/2; 48/3; 64/3; 96/3; 163/3  
C;Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match 74.5%; Score 660; DB 2; Length 205;  
Best Local Similarity 66.2%; Pred. No. 3.2e-46;  
Matches 137; Conservative 16; Mismatches 18; Indels 36; Gaps 4;

QY 1 MSSMNPEDYLFKLLIGDSGVGKSCLLLRPADDTYESYISTIGVDFKIRTIELDGKTI 60  
DB 1 MAANPEYDLFKLLIGDSGVGKSCLLLRPADDTYESYISTIGVDFKIRTIELDGKTI 60  
QY 61 KLOI-----ESFNNKQWLOEIDRYASENNKL 88  
DB 61 KLOIMDTAGGERFTITSSYRGAGIIVYDITDQETFNNKQWLOEIDRYACENYKL 120  
QY 89 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGP 148  
DB 121 LVGNKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEOSFMTAAEIKRMGPVQ 180  
QY 149 TAGGAESKNVKI-QSTPVKQSGGCC 173  
DB 181 GAGCA--PGVATIGSPVQDKKSGGCC 205

RESULT 10  
S30096  
GTP-binding protein ypt1 [similarity] - Neurospora crassa  
N;Alternate names: protein B910.240; ras-related protein ypt1  
C;Species: Neurospora crassa  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 19-Jan-2001  
C;Accession: S30096; T49787  
R;Heintz, K.; Palme, K.; Diefenthal, T.; Russo, V.E.A.  
Mol. Gen. Genet. 235, 413-421, 1992  
A;Title: The Ncyp1 gene from Neurospora crassa is located on chromosome 2: molecular  
A;Reference number: S30096; MUID:93101148; PMID:1361212  
A;Accession: S30096  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-203 <HEI>  
A;Cross-references: GB:S51252; NID:g262045; PIDN:AAB24564.1; PID:g262046  
R;Schulte, U.; Aign, V.; Hehlsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatu  
submitted to the Protein Sequence Database, May 2000  
A;Reference number: Z25022  
A;Accession: T49787  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-203 <SCH>

A:Cross-references: EMBL:AL356324; GSPDB:GN00116; NCSPB:B9J10.240  
 A:Experimental source: BAC clone B9J10; strain OR4A  
 C:Genetics:  
 A:Gene: ypt1; NCSPB:B9J10.240  
 A:Map position: 6  
 A:Introns: 4/2; 45/3; 137/3  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; lipoprotein; nucleotide binding; P-loop; prenylated cysteine  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKAD motif  
 F:151-153/Region: GTP-binding SAK/L motif  
 F:21,22,40,121,122,124,131/Binding site: Mg-GTP (lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
 F:202,203/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 74.3%; Score 658.5; DB 2; Length 203;  
 Best Local Similarity 66.5%; Pred. No. 4,2e-46;  
 Matches 135; Conservative 11; Mismatches 24; Indels 33; Gaps 2;

QY 4 MNPEDYLFKLLIGDSGVGKSCLLRFADDTYTESYSTIGVDFKIRTIELDKRTIKLQ 63  
 Db 1 MNPEDYLFKLLIGDSGVGKSCLLRFADDTYTESYSTIGVDFKIRTIELDKRTIKLQ 60  
 QY 64 I-----ESFNNVQWLOEIDRYASENNKLLVG 91  
 Db 61 IMDTAGQERFRTITSSYRGAGICVYDVTDMSFNNVQWLOEIDRYATEGVNKLIVG 120  
 QY 92 NKCDLITKKVYDITAKEFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGATAG 151  
 Db 121 NKSMTKRVKVEYTAKEFADSLGIPLETSAKNSVNEOAFITMARQIKERMSSIATN 180  
 QY 152 GAERKSNVKI-OSTPVKOSGGGCC 173  
 Db 181 NTKASVNVSPGRCVGNSSGGCC 203

RESULT 11  
 JCI4105  
 GTP-binding protein yptC1 - Chlamydomonas reinhardtii  
 N:Alternate names: membrane vesicle transport protein yptC1; ras-like yptC1 protein; sma  
 C:Species: Chlamydomonas reinhardtii  
 C:Date: 21-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 02-Feb-2001  
 C:Accession: JCI4105  
 R:Diemater: W.; Fabry, S.; Huber, H.; Schmitt, R.  
 Gene 158, 41-50, 1995  
 A:Title: Analysis of a family of ypt genes and their products from Chlamydomonas reinhar  
 A:Reference number: JCI4105; MUID:95309723; PMID:7789609  
 A:Accession: JCI4105  
 A:Molecule type: mRNA  
 A:Residues: 1-203 <CDS>  
 A:Cross-references: GB:U13168; NID:9806723; PIDN:AA82727.1; PID:9806724  
 C:Comments: This protein plays an essential role in the regulation of intracellular membr  
 C:Genetics:  
 A:Gene: yptC1  
 A:Introns: 5/1; 10/2; 29/2; 45/2; 61/2; 85/2; 137/2; 172/1  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKAD motif  
 F:151-153/Region: GTP-binding SAK/L motif  
 F:21,22,40,121,122,124,131/Binding site: Mg-GTP (lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta  
 F:202,203/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 73.0%; Score 646.5; DB 2; Length 203;  
 Best Local Similarity 65.7%; Pred. No. 3.9e-45;  
 Matches 134; Conservative 9; Mismatches 26; Indels 35; Gaps 3;

QY 4 MNPEDYLFKLLIGDSGVGKSCLLRFADDTYTESYSTIGVDFKIRTIELDKRTIKLQ 63  
 Db 1 MNPEDYLFKLLIGDSGVGKSCLLRFADDTYTESYSTIGVDFKIRTIELDKRTIKLQ 60

QY 64 I-----ESFNNVQWLOEIDRYASENNKLLVG 91  
 Db 61 IMDTAGQERFRTITSSYRGAGICVYDVTDDESFNNVQWLOEIDRYASENNKLLVG 120  
 QY 92 NKCDLITKKVYDITAKEFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGATAG 149  
 Db 121 NKSMTKRVKVEYTAKEFADSLGIPLETSAKNTNVEOAFITMARQIKERMSSIATN 180  
 QY 150 AGAERKSNVKI-OSTPVKOSGGGCC 173  
 Db 181 KAGGVPVBPQ-EGKPIINSSGCC 203

RESULT 12  
 T50323  
 ypt1-related protein 1 [imported] - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 02-Sep-2000  
 C:Accession: T50323  
 R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaure, V.; Gilbert,  
 submitted to the EMBL Data Library, January 2000  
 A:Reference number: 225061  
 A:Accession: T50323  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-203 <MCD>  
 A:Cross-references: EMBL:AL136336; PIDN:CAB6454.1; GSPDB:GN00067; SPDB:SPBC1703.10  
 C:Genetics:  
 A:Gene: SPDB:SPBC1703.10  
 A:Map position: 2  
 A:Introns: 4/2; 45/3; 61/3; 94/1  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 F:9-124/Domain: translation elongation factor Tu homology <ETU>

Query Match 73.0%; Score 646.5; DB 2; Length 203;  
 Best Local Similarity 64.5%; Pred. No. 3.9e-45;  
 Matches 131; Conservative 16; Mismatches 23; Indels 33; Gaps 2;

QY 4 MNPEDYLFKLLIGDSGVGKSCLLRFADDTYTESYSTIGVDFKIRTIELDKRTIKLQ 63  
 Db 1 MNPEDYLFKLLIGDSGVGKSCLLRFADDTYTESYSTIGVDFKIRTIELDKRTIKLQ 60  
 QY 64 I-----ESFNNVQWLOEIDRYASENNKLLVG 91  
 Db 61 IMDTAGQERFRTITSSYRGAGICVYDVTDDESFNNVQWLOEIDRYATEGVNKLIVG 120  
 QY 92 NKCDLITKKVYDITAKEFADSLGIPLETSAKNTNVEOSFMTMAEIKRMGPGATAG 151  
 Db 121 NKSMTKRVKVEYTAKEFADSLGIPLETSAKNTNVEOAFITMARQIKERMSSIATN 180  
 QY 152 GAERKSNVKI-OSTPVKOSGGGCC 173  
 Db 181 SNAKSVKVGQGTNVSSQSSNCC 203

RESULT 13  
 JCI247  
 GTP-binding protein yptV1 - Volvox carterii  
 C:Species: Volvox carterii  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 02-Feb-2001  
 C:Accession: JCI247  
 R:Fabry, S.; Netz, N.; Huber, H.; Palme, K.; Jaenicke, L.; Schmitt, R.  
 Gene 118, 153-162, 1992  
 A:Title: The yptV1 gene encodes a small G-protein in the green alga Volvox carterii: G  
 A:Reference number: JCI247; MUID:92380499; PMID:1511889  
 A:Accession: JCI247  
 A:Molecule type: DNA  
 A:Residues: 1-203 <FAB>  
 A:Cross-references: GB:M93438; NID:g170660; PIDN:AAA34255.1; PID:g170661  
 C:Genetics:  
 A:Gene: yptV1  
 A:Introns: 5/2; 10/3; 29/3; 45/3; 61/3; 85/3; 137/3; 172/2

C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; lipo-protein; membrane trafficking; nucleotide binding; P-loop;  
 F:9-14/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:40-45/Region: GTP binding #status predicted  
 F:61-68/Region: GTP binding #status predicted  
 F:120-126/Region: GTP binding #status predicted  
 F:147-153/Region: GTP binding #status predicted  
 F:202,203/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match 72.5%; Score 642.5; DB 2; Length 203;  
 Best Local Similarity 65.0%; Pred. No. 8.2e-45;  
 Matches 134; Conservative 6; Mismatches 27; Indels 39; Gaps 3;

QY 4 MNPEYDLFFLLIGSGVGKSCLLRFADDTYESYSTIGVDFKIRITELDGKTIKQ 63  
 DB 1 MNPEYDLFFLLIGSGVGKSCLLRFADDTYESYSTIGVDFKIRITELDGKTIKQ 60  
 QY 64 I-----ESFNNVKQMLQIEDRYASENVNKLIVG 91  
 DB 61 IMDTAGQERERTTSSYRGAGHIIIVYDVTQDESFNNVKQMLQIEDRYASENVNKLIVG 120  
 QY 92 NKCDLTKKVVDTTAKFADSDIGIPLETSAKNATVDEQSFMTMAEIKRKG----PG 147  
 DB 121 NKSDLTGKKVVDYQAKAFADDEIGIPLETSAKNATVDEQAFMTMAEIKRNASQPVPP 180  
 QY 148 ATAGAERSNVKIQSTPVKQSGGCC 173  
 DB 181 KPGGPVVRPT---EGKRINNKSSSC 203

## RESULT 14

JC5337  
 GTP-binding protein ypt1 - Phytophthora infestans  
 N:Alternate names: monomeric GTP-binding protein  
 C:Species: Phytophthora infestans (potato late blight agent)  
 C>Date: 14-May-1997 #sequence\_revision 18-Jul-1997 #text\_change 02-Feb-2001  
 R:Chen, Y.; Roxy, R.  
 Gene 181, 89-94, 1996  
 A:Title: Characterization of a Phytophthora infestans gene involved in vesicle transport  
 A:Reference number: JC5337; MID:97128773; PMID:8973313  
 A:Accession: JC5337  
 A:Molecule type: DNA  
 A:Residues: 1-201 <CHE>  
 A:Cross-references: GB:U30474; NID:940431; PIDN:ABA0355.1; PID:940432  
 C:Comment: This protein is involved in vesicle transport between the endoplasmic reticulum and Golgi apparatus.

A:Gene: ypt1  
 A:Introns: 5/2; 45/3; 61/3; 108/2  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop  
 F:9-14/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKXD motif  
 F:151-153/Region: GTP-binding SAK/L motif  
 F:21,22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #status predicted

Query Match 71.3%; Score 631.5; DB 2; Length 201;  
 Best Local Similarity 67.5%; Pred. No. 6.2e-44;  
 Matches 139; Conservative 8; Mismatches 18; Indels 41; Gaps 5;

QY 4 MNPEYDLFFLLIGSGVGKSCLLRFADDTYESYSTIGVDFKIRITELDGKTIKQ 63  
 DB 1 MNPEYDLFFLLIGSGVGKSCLLRFADDTYESYSTIGVDFKIRITELDGKTIKQ 60  
 QY 64 I-----ESFNNVKQMLQIEDRYASENVNKLIVG 91  
 DB 61 IMDTAGQERERTTSSYRGAGHIIIVYDVTQDESFNNVKQMLQIEDRYASENVNKLIVG 120  
 QY 92 NKCDLTKKVVDTTAKFADSDIGIPLETSAKNATVDEQSFMTMAEIKRKGPGATAG 151  
 DB 121 NKSDLTGKKVVDYQAKAFADDEIGIPLETSAKNATVDEQAFMTMAEIKRKGPGATAG 177

QY 152 GAERSNVKIQSTPVKQ--SGGG--CC 173  
 DB 178 VAPKAGVKL--TGGQGVPSNGSKCC 201

## RESULT 15

S34253  
 GTP-binding protein, ras-related - common tobacco  
 C:Species: Nicotiana tabacum (common tobacco)  
 C>Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 02-Feb-2001  
 C:Accession: S34253  
 R:Napier, J.A.; Shewry, P.R.  
 Submitted to the EMBL Data Library, May 1993  
 A:Reference number: S34253  
 A:Accession: S34253  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-203 <NAP>  
 A:Cross-references: EMBL:X72212; NID:g311906; PIDN:CAA51011.1; PID:g311907  
 C:Superfamily: ras transforming protein; translation elongation factor Tu homology  
 C:Keywords: GTP binding; nucleotide binding; P-loop  
 F:9-14/Domain: translation elongation factor Tu homology <ETU>  
 F:15-22/Region: nucleotide-binding motif A (P-loop)  
 F:121-124/Region: GTP-binding NKXD motif  
 F:151-153/Region: GTP-binding SAK/L motif  
 F:21,22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #status predicted

Query Match 70.8%; Score 627.5; DB 2; Length 203;  
 Best Local Similarity 63.9%; Pred. No. 1.3e-43;  
 Matches 129; Conservative 10; Mismatches 30; Indels 33; Gaps 2;

QY 4 MNPEYDLFFLLIGSGVGKSCLLRFADDTYESYSTIGVDFKIRITELDGKTIKQ 63  
 DB 1 MNPEYDLFFLLIGSGVGKSCLLRFADDTYESYSTIGVDFKIRITELDGKTIKQ 60  
 QY 64 I-----ESFNNVKQMLQIEDRYASENVNKLIVG 91  
 DB 61 IMDTAGQERERTTSSYRGAGHIIIVYDVTQDESFNNVKQMLQIEDRYASENVNKLIVG 120  
 QY 92 NKCDLTKKVVDTTAKFADSDIGIPLETSAKNATVDEQSFMTMAEIKRKGPGATAG 151  
 DB 121 NKSDLTGKKVVDYQAKAFADDEIGIPLETSAKNATVDEQAFMTMAEIKRNASQPVPP 180  
 QY 152 GAERSNVKIQSTPVKQSGGCC 173  
 DB 181 SAKPPTVNINGQVYTGQ--GGCC 201

Search completed: November 17, 2002, 13:07:56  
 Job time : 44 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 17, 2002, 11:42:42 : Search time 25 Seconds  
(without alignments)  
287.016 Million cell updates/sec

Title: US-09-820-003a-2

Perfect score: 886  
Sequence: 1 MSNMPEYDYLFKLLIGDS.....EKSNKIOSTEVKSGGCC 173

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	860	97.1	205	1 RB1A_HUMAN	P11476 homo sapien
2	849	95.8	205	1 RB1A_RAT	P05711 rattus norv
3	828	93.5	202	1 RB1L_D1SOM	P22123 discopysg o
4	760.5	85.8	201	1 RB1B_RAT	P10533 rattus norv
5	732	82.7	203	1 RB1L_MOUSE	O03974 lymnaea sta
6	658.2	74.3	203	1 YPT1_MOUSE	P33723 neosporea
7	646.5	73.0	203	1 YPT1_CHLRE	O39571 chlamydomon
8	646.5	73.0	203	1 YPT1_CHLRE	P11620 schizosacch
9	642.5	72.5	203	1 YPT1_VOLCA	P31584 volvox cart
10	631.5	71.3	201	1 YPT1_PHYTN	P01890 phytophthor
11	619.5	69.9	258	1 ARAB_ARATH	O01880 arabidopsis
12	618.5	69.8	203	1 YPT2_MAIZE	O05737 zea mays (m
13	613.5	69.2	202	1 RIC1_YEAST	P40392 oryza sativ
14	571	64.4	206	1 YPT1_YEAST	P01133 saccharomyc
15	542	61.2	199	1 RB1A_D1CIDI	P34139 dictyostell
16	538	59.0	208	1 YPT1_MAIZE	P16976 zea mays (m
17	523	52.9	208	1 SAS1_D1CIDI	P20790 dictyostell
18	468.5	52.0	203	1 SAS2_D1CIDI	P20791 dictyostell
19	461	49.9	215	1 RB1L_BETUV	O39433 beta vulgar
20	442	49.5	216	1 ARAB_ARATH	P28186 arabidopsis
21	438.5	49.5	200	1 YPT2_SCHPO	P17609 schizosacch
22	435.5	47.9	210	1 RB1B_D1SOM	P22128 discopysg o
23	424	47.2	207	1 RB1B_HUMAN	P24407 homo sapien
24	418	46.8	199	1 RB1B_RAT	P70550 rattus norv
25	417	46.7	207	1 RB1B_D1CIDI	O92930 homo sapien
26	414.5	46.3	206	1 YPT2_VOLCA	P31411 dictyostell
27	410.5	45.8	217	1 RB1L_MOUSE	P55258 mus musculu
28	405.5	45.5	200	1 RB1L_HUMAN	P31861 volvox cart
29	403.5	45.4	200	1 RB10_CANFA	O88386 canis famli
30	402.5	44.9	200	1 RAO1_D1SOM	P21217 discopysg o
31	398	44.9	203	1 RYL1_YARLI	P41924 yarrowia li
32	398	44.9	203	1 RYL1_YARLI	P41924 yarrowia li
33	398	44.9	203	1 RYL1_YARLI	P41924 yarrowia li

34	384.5	43.4	203	1 RB13_HUMAN	P51153 homo sapien
35	374.5	42.3	200	1 RB10_RAT	P35281 rattus norv
36	373.5	42.2	215	1 SEC4_YEAST	P07560 saccharomyc
37	360.5	40.7	201	1 RB35_HUMAN	O15286 homo sapien
38	360.5	40.6	210	1 SEC4_CANAL	O14462 candida alb
39	352	39.7	209	1 RB2A_MAIZE	P49104 zea mays (m
40	349.5	38.9	210	1 RB2B_MAIZE	P49104 zea mays (m
41	344.5	38.9	213	1 YPT4_CHLRE	O39570 chlamydomon
42	342	38.6	219	1 RB3A_BOVIN	P11023 bos taurus
43	339	38.3	213	1 YPT4_VOLCA	P36863 volvox cart
44	336	37.9	220	1 RB3_DROME	P25228 drosophila
45	336	37.9	220	1 RB3A_HUMAN	P20336 homo sapien

## ALIGNMENTS

RESULT 1  
RB1A\_HUMAN STANDARD; PRT; 205 AA.  
ID RB1A\_HUMAN  
AC P11476:  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Ras-related protein Rab-1A (YPT1-related protein).  
GN RAB1 OR RAB1A.  
OS Homo sapiens (Human), and  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_Taxid=9606, 10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Human: Pubmed-2501306.  
RX MEDLINE-89308668; Pubmed-2501306.  
RA Zahraoui A., Touchot N., Chardin P., Tavittian A.;  
PT "The human Rab genes encode a family of GTP-binding proteins related  
to yeast YPT1 and SEC4 products involved in secretion.";  
J Biol. Chem. 264:12394-12401(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Human: TISSUE-Placenta;  
RX Strausberg R.;  
PT Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Mouse;  
RX MEDLINE-88166649; Pubmed-3127202;  
RA Haubruck H., Disela C., Wagner P., Gallwitz D.;  
PT "The ras-related ypt protein is an ubiquitous eukaryotic protein:  
isolation and sequence analysis of mouse cDNA clones highly  
homologous to the yeast YPT1 gene.";  
EMBO J. 6:4049-4053(1987).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Mouse;  
RX MEDLINE-89386011; Pubmed-2506528;  
RA Wichmann H., Disela C., Haubruck H., Gallwitz D.;  
PT "Nucleotide sequence of the mouse ypt1 gene encoding a ras-related  
GTP-binding protein.";  
Nucleic Acids Res. 17:6737-6738(1989).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES-Mouse;  
RX Wu G., Dorn G.W., II;  
PT "Mouse Rab1a, member of Ras oncogene family mRNA.";  
Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP PHOSPHORYLATION BY CDC2.  
RX MEDLINE-91218852; Pubmed-1902553;  
RA Bailey E., McCreffrey M., Touchot N., Zahraoui A., Goud B., Bornens M.;  
PT "Phosphorylation of two small GTP-binding proteins of the Rab family  
by p34cdc2.";

RL Nature 350:715-718(1991).  
 RN [7]  
 RP ISOPRENOLID.  
 RX MEDLINE-91296801; PubMed=1648736;  
 RA Khosravi-Far R., Lutz R.J., Cox A.D., Conroy L., Bourne J.R.,  
 RT Shiensky M., Balch W.E., Buss J.E., Der C.J.;  
 "Isoprenoid modification of rab proteins terminating in CC or CXC  
 motifs.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6264-6268(1991).  
 CC -1- FUNCTION: PROBABLY REQUIRED FOR TRANSIT OF PROTEIN FROM THE ER  
 CC THROUGH GOLGI COMPARTMENT.  
 CC -1- SUBCELLULAR LOCATION: GOLGI COMPARTMENT.  
 CC -1- PTM: PHOSPHORYLATED BY CDC2-KINASE DURING MITOSIS.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M28209; AAA60240.1; -;  
 DR EMBL: BC000905; AAA00905.1; -;  
 DR EMBL: Y00094; CAA68284.1; -;  
 DR EMBL: X15744; CAA33760.1; -;  
 DR EMBL: X15745; CAA33760.1; JOINED.  
 DR EMBL: X15746; CAA33760.1; JOINED.  
 DR EMBL: X15747; CAA33760.1; JOINED.  
 DR EMBL: AF226873; AAF33844.1; -;  
 DR PIR: S05551; TVMSIP.  
 DR PIR: A34323; A34323.  
 DR PIR: S19104; S19104.  
 DR HSSP: P05713; 3RAB.  
 DR PMAA-2DPAGE: P11476; -;  
 DR Genew: H9NC; 9758; RAB1A.  
 DR MIM: 179508; -;  
 DR MGD: MGI:97842; Rab1.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trasfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; Ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfams: TIGR00231; small\_gtp; 1.  
 DR GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack;  
 KM Endoplasmic reticulum; phosphorylation.  
 FT NP\_BIND 18 25 GTP (BY SIMILARITY).  
 FT NP\_BIND 66 70 GTP (BY SIMILARITY).  
 FT DOMAIN 40 48 EFFECTOR REGION (BY SIMILARITY).  
 FT MOD\_RES 194 194 PHOSPHORYLATION (BY CDC2) (PROBABLE).  
 FT LIPID 204 204 GERANYL-GERANYL.  
 FT LIPID 205 205 GERANYL-GERANYL.  
 SQ SEQUENCE 205 AA; 22678 MW; B2ABF4E3B0FB1D6 CRC64;

Query Match 97.1%; Score 860; DB 1; Length 205;  
 Best local Similarity 84.4%; Pred. No. 9.3e-65;  
 Matches 173; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

QY 1 MSSMNPEDYLFKLLIGDSGVGKSCLLFRPADTYESYISTIGVDFKRTIELDKTI 60  
 DB 1 MSSMNPEDYLFKLLIGDSGVGKSCLLFRPADTYESYISTIGVDFKRTIELDKTI 60  
 QY 61 KLOI-----ESFNWKQWLOEIDRYASENNKL 88  
 DB 61 KLOIWDAGGERFRTTISYRGAGHITVYDVYDQESFNWKQWLOEIDRYASENNKL 120  
 QY 89 LVGNKCDLTTRKVVDTTAKFADSLGIPLETSAKNATNVEOSFMMAAEIKRKGPGA 148  
 DB 121 LVGNKCDLTTRKVVDTTAKFADSLGIPLETSAKNATNVEOSFMMAAEIKRKGPGA 180

QY 149 TAGAERKSNVKIOSTPVKOSGGGCC 173  
 DB 181 TAGAERKSNVKIOSTPVKOSGGGCC 205

RESULT 2  
 ID RBA1\_RAT STANDARD; PRT; 205 AA.  
 AC P05711;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ras-related protein Rab-1A.  
 OS RAB1 OR RAB1A.  
 GN Rattus norvegicus (Rat), and  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116, 9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Rat;  
 RX MEDLINE-88068563; PubMed=3317403;  
 RA Touchot N., Chardin P., Tavittian A.;  
 RT "Four additional members of the ras gene superfamily isolated by an  
 RT oligonucleotide strategy: molecular cloning of YPT-related cDNAs from  
 RT a rat brain library.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8210-8214(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-C.familiaris; STRAIN-Cocker spaniel;  
 RX MEDLINE-91061765; PubMed=2123294;  
 RA Chavrier P., Vignon M., Sander C., Simons K., Zerial M.;  
 RT "Molecular cloning of YPT/SEC4-related cDNAs from an epithelial cell  
 RT line.";  
 RL Mol. Cell. Biol. 10:6578-6585(1990).  
 RN [3]  
 RP ISOPRENOLID.  
 RX MEDLINE-91296801; PubMed=1648736;  
 RA Khosravi-Far R., Lutz R.J., Cox A.D., Conroy L., Bourne J.R.,  
 RA Shiensky M., Balch W.E., Buss J.E., Der C.J.;  
 RT "Isoprenoid modification of rab proteins terminating in CC or CXC  
 RT motifs.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:6264-6268(1991).  
 CC -1- FUNCTION: PROBABLY REQUIRED FOR TRANSIT OF PROTEIN FROM THE ER  
 CC THROUGH GOLGI COMPARTMENT.  
 CC -1- MISCELLANEOUS: RAB-1A BINDS GTP AND GDP AND POSSESS INTRINSIC  
 CC GTPASE ACTIVITY.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: J02998; AAA42006.1; -;  
 DR EMBL: X56384; CAB56775.1; ALT-SEQ.  
 DR PIR: A36364; A36364.  
 DR PIR: A39963; A39963.  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trasfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; Ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfams: TIGR00231; small\_gtp; 1.  
 DR GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack;  
 KM Endoplasmic reticulum.  
 FT NP\_BIND 18 25 GTP (BY SIMILARITY).



```

FT NP_BIND 66 70 GTP (BY SIMILARITY).
FT NP_BIND 124 127 GTP (BY SIMILARITY).
FT DOMAIN 40 48 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 204 204 GERANYL-GERANYL.
FT LIPID 205 205 GERANYL-GERANYL.
SQ SEQUENCE 205 AA: 22763 MW: 34649523BDF0DEB CRC64:

Query Match 95.8%; Score 849; DB 1; Length 205;
Best Local Similarity 83.4%; Pred. No. 7.7e-64;
Matches 171; Conservative 0; Mismatches 2; Indels 32; Gaps 1;

QY 1 MSNMPEDYFLKLLIGDSGVKSCLLRPADDTYESTISTGVDFKRTIELDKTI 60
   1 MSNMPEDYFLKLLIGDSGVKSCLLRPADDTYESTISTGVDFKRTIELDKTI 60
DB 1 KIQIMDTAGQERRITTSYRGAGIIVYDVTDQSEFNKQWLOEIDRVASENVKL 120
   61 KIQIMDTAGQERRITTSYRGAGIIVYDVTDQSEFNKQWLOEIDRVASENVKL 120
QY 61 KIQT-----ESFNKQWLOEIDRVASENVKL 88
   61 KIQIMDTAGQERRITTSYRGAGIIVYDVTDQSEFNKQWLOEIDRVASENVKL 120
DB 61 KIQIMDTAGQERRITTSYRGAGIIVYDVTDQSEFNKQWLOEIDRVASENVKL 120
   89 LVNKKDLTKKVVDTTAAKEFADSLGIPLETSAKNATNVESEFMTMAEIKRRMGPGA 148
QY 121 LVNKKDLTKKVVDTTAAKEFADSLGIPLETSAKNATNVESEFMTMAEIKRRMGPGA 180
   121 LVNKKDLTKKVVDTTAAKEFADSLGIPLETSAKNATNVESEFMTMAEIKRRMGPGA 180
QY 149 TAGGAESNVKIQSTPVKQSGGCC 173
   149 TAGGAESNVKIQSTPVKQSGGCC 173
DB 181 TAGGAESNVKIQSTPVKQSGGCC 205
   181 TAGGAESNVKIQSTPVKQSGGCC 205

RESULT 3
RAB1_DISOM STANDARD; PRT; 202 AA.
AC P22125;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Ras-related protein ORAB-1.
OS Discopyle ommata (Electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalea; Hypnosqualea; Pisticoralea; Batoidae;
OC Torpediniformes; Narcinoidae; Narciniidae; Discopyge.
OX NCBI_TaxID=7785;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Electric lobe;
RX MEDLINE=91115900; PubMed=1899244;
RA Nysse J.K., Elferink L.A., Scheller R.H.;
RT "A family of ras-like GTP-binding proteins expressed in electromotor
   neurons."
RL J. Biol. Chem. 266:2675-2680(1991).
CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
   TRAFFIC (BY SIMILARITY).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
DR EMBL: M38393; AAA9234.1; .
DR F1R: D38625; D38625.
DR HSP: P05713; 3RAB.
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001806; Ras_Transfmg.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00071; Ras_1.
DR PRINTS: PR00449; RASTRNSFRNG.
DR SMART: SM00175; RAB; 1.
DR TIGRFAMs: TIGR00231; small_GTP; 1.
KM GTP-binding; Lipoprotein; Prenylation; Protein transport.
FT NP_BIND 15 22 GTP (BY SIMILARITY).

```

```

FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).
FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).
FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).
SQ SEQUENCE 202 AA: 22333 MW: 31078502BCDD1B9 CRC64:

Query Match 93.5%; Score 828; DB 1; Length 202;
Best Local Similarity 81.7%; Pred. No. 4.2e-62;
Matches 165; Conservative 3; Mismatches 2; Indels 32; Gaps 1;

QY 4 MNPEDYFLKLLIGDSGVKSCLLRPADDTYESTISTGVDFKRTIELDKTIKQ 63
   1 MNPEDYFLKLLIGDSGVKSCLLRPADDTYESTISTGVDFKRTIELDKTIKQ 60
DB 1 MNPEDYFLKLLIGDSGVKSCLLRPADDTYESTISTGVDFKRTIELDKTIKQ 60
   64 I-----ESFNKQWLOEIDRVASENVKL 91
DB 61 IMPTAQERFRTITSSYRGAGIIVYDVTDQSEFNKQWLOEIDRVASENVKL 120
   92 NKCDLTKKVVDTTAAKEFADSLGIPLETSAKNATNVESEFMTMAEIKRRMGPGATAG 151
QY 121 NKCDLTKKVVDTTAAKEFADSLGIPLETSAKNATNVESEFMTMAEIKRRMGPGATAG 180
   121 NKCDLTKKVVDTTAAKEFADSLGIPLETSAKNATNVESEFMTMAEIKRRMGPGATAG 180
QY 152 GAESNVKIQSTPVKQSGGCC 173
   152 GAESNVKIQSTPVKQSGGCC 173
DB 181 GSEKSNVNIQSTPVKQSGGCC 202
   181 GSEKSNVNIQSTPVKQSGGCC 202

RESULT 4
RAB1B_RAT STANDARD; PRT; 201 AA.
AC P10536;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Ras-related protein Rab-1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89160341; PubMed=2493636;
RA Zahraoui A., Touchot N., Charidin P., Tavittian A.;
RT "Nucleotide sequence of a rat cDNA: rab1b, encoding a rab1-YPT
   related protein."
RL Nucleic Acids Res. 17:1770-1770(1989).
CC CHARACTERIZATION.
RX MEDLINE=9003316; PubMed=2509243;
RA Touchot N., Zahraoui A., Vleth E., Tavittian A.;
RT "Biochemical properties of the YPT-related rab1b protein. Comparison
   with rab1A."
RL FEBS Lett. 256:79-84(1989).
RN [3]
RP ISOPRENOID.
RX MEDLINE=91296801; PubMed=1648736;
RA Khoshdeli-Far R., Lutz R.J., Cox A.D., Conroy L., Bourne J.R.,
RA Shensky M., Balch W.E., Buss J.E., Der C.J.;
RT "Isoprenoid modification of rab proteins terminating in CC or CXC
   motifs."
RL Proc. Natl. Acad. Sci. U.S.A. 88:6264-6268(1991).
CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR
   TRAFFIC (BY SIMILARITY).
CC -1- MISCELLANEOUS: RAB-1B BINDS GTP AND GDP AND POSSESS INTRINSIC
   GTPASE ACTIVITY.
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X13905; CAA32105.1; -.  
DR PIR: S03189; S03189.  
DR PIR: S06147; S06147.  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsfmg.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; ras; 1.  
DR PRINTS: PR00449; RASTRNSFRMNG.  
DR SMART: SM00175; Rab; 1.  
DR TIGRFRAMS: TIGR00231; small\_gtp; 1.  
KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
FT DOMAIN 37 45 EFFECTOR REGION (BY SIMILARITY).  
FT LIPID 200 200 GERANYL-GERANYL.  
FT LIPID 201 201 GERANYL-GERANYL.  
FT MUTAGEN 21 21 K->M: ABOLISHES GTP-BINDING.  
SQ SEQUENCE 201 AA; 22163 MW; 8D3EDC2AEF4A2PE CRC64;  
  
Query Match 85.8%; Score 760.5; DB 1; Length 201;  
Best Local Similarity 76.2%; Pred. No. 1,7e-56;  
Matches 154; Conservative 7; Mismatches 8; Indels 33; Gaps 2;  
  
QY 4 MNPEYDLFKLLIGDSGVKSCLLRFADDTYESYISTGVDFKRTIELDKTI 63  
DB 1 MNPEYDLFKLLIGDSGVKSCLLRFADDTYESYISTGVDFKRTIELDKTI 60  
QY 64 I-----ESFNNVKOMLOEIDRYASENVNKL 91  
DB 61 IMDPAGGERRTVTSYRGAGHGIIVYDVTQDSYANVKOMLOEIDRYASENVNKL 120  
QY 92 NKCDLTKRKVVDTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPATAG 151  
DB 121 NKSDLTKRKVVDTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGPATAG 180  
QY 152 GAESKNVKIOSTPYKOSGGCC 173  
DB 181 G-ERPNIKIDSTPYKASAGGCC 201  
  
RESULT 5  
RABL\_LYMST  
ID RABL\_LYMST STANDARD; PRT; 205 AA.  
AC 005974;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE Ras-related protein Rab-1A.  
GN RAB1A.  
OS Lymanaea stagnalis (Great pond snail).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
OC Lymnaeidae; Lymnaea.  
OX NCBI\_TaxID=6523;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94039042; PubMed=8223561;  
RA Agterberg M., van Die I., Yang H., Andriessen J.A.,  
van Tetering A., van den Elinden D.H., Ploegh H.L.;  
RT "Isolation and characterization of three cDNAs coding for Rab  
proteins from the alveolar gland of the mollusc Lymanaea stagnalis.";  
RL Eur. J. Biochem. 217:241-246(1993).  
CC -1- FUNCTION: PROBABLY REQUIRED FOR TRANSIT OF PROTEIN FROM THE ER  
THROUGH GOLGI COMPARTMENT (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: GOLGI COMPARTMENT (BY SIMILARITY).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
4 5 6

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X72688; CAA51233.1; -.  
DR PIR: S32206; S32206.  
DR PIR: S38339; S38339.  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsfmg.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; ras; 1.  
DR PRINTS: PR00449; RASTRNSFRMNG.  
DR SMART: SM00175; Rab; 1.  
DR TIGRFRAMS: TIGR00231; small\_gtp; 1.  
KW GTP-binding; Lipoprotein; Prenylation; Protein transport; Golgi stack;  
KW Endoplasmic reticulum.  
FT NP\_BIND 18 25 GTP (BY SIMILARITY).  
FT NP\_BIND 66 70 GTP (BY SIMILARITY).  
FT NP\_BIND 124 127 GTP (BY SIMILARITY).  
FT DOMAIN 40 48 EFFECTOR REGION (BY SIMILARITY).  
FT LIPID 204 204 GERANYL-GERANYL (BY SIMILARITY).  
FT LIPID 205 205 GERANYL-GERANYL (BY SIMILARITY).  
SQ SEQUENCE 205 AA; 22760 MW; D5E0BDDCBA426FF CRC64;  
  
Query Match 82.7%; Score 733; DB 1; Length 205;  
Best Local Similarity 73.8%; Pred. No. 3.4e-54;  
Matches 152; Conservative 7; Mismatches 13; Indels 34; Gaps 3;  
  
QY 1 MSSNPEYDLFKLLIGDSGVKSCLLRFADDTYESYISTGVDFKRTIELDKTI 60  
DB 1 MSTNMPDYDLFKLLIGDSGVKSCLLRFADDTYESYISTGVDFKRTIELDKTI 60  
QY 61 KLOI-----ESFNNVKOMLOEIDRYASENVNKL 88  
DB 61 KLOI-----ESFNNVKOMLOEIDRYASENVNKL 120  
QY 89 LVGNKCDLTKRKVVDTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGP 148  
DB 121 LVGNKCDLTKRKVVDTTAKAFADSLGIPLETSAKNATNEQSFMTMAEIKRKMGP-1 179  
QY 149 TAGGAESKNVKI-OSTPYKOSGGCC 173  
DB 180 TAASDSKPSVKINSSTPYKASAGGCC 205  
  
RESULT 6  
YPTL\_NEUCR  
ID YPTL\_NEUCR STANDARD; PRT; 203 AA.  
AC P33723;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE GTP-binding protein ypt1.  
GN YPT-1 OR B9J10.240.  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93101148; PubMed=1361212;  
RA Heintz K., Palme K., Diefenthal F., Russo V.E.A.;  
RT "The Ncyp1 gene from Neurospora crassa is located on chromosome 2:  
RT molecular cloning and structural analysis.";  
RL Mol. Gen. Genet. 235:413-421(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=74-OR23-1A;  
RA Schulte U., Algen V., Hohelsel J., Brandt P., Fartmann B., Holland R.,

RA Nyakatura G., Mewes H.-W., Mannhaupt G.;  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S51252; AAB24564.1; -  
 DR EMBL: AL356324; CAB92031.1; -  
 DR PIR: S30096; S30096.  
 DR HSSP: F01112; 1PLU.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfrrng.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAms: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NP-BIND 15 22 GTP (BY SIMILARITY).  
 FT NP-BIND 63 67 GTP (BY SIMILARITY).  
 FT NP-BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 202 202 GERANYL-GERANYL.  
 FT LIPID 203 203 GERANYL-GERANYL.  
 FT LIPID 203 203 GERANYL-GERANYL.  
 SQ SEQUENCE 203 AA; 22476 MW; 20F5A1AB3400EFLD CRC64;  
 Query Match 74.3%; Score 658.5; DB 1; Length 203;  
 Best Local Similarity 66.5%; Pred. No. 5.3e-48;  
 Matches 135; Conservative 11; Mismatches 24; Indels 33; Gaps 2;  
 QY 4 MNPEVDYLFKLLIGDSGVKSCLLLRPADDTYESYSTIGVDFKIRTELDGKTIKQ 63  
 DB 1 MNPEVDYLFKLLIGDSGVKSCLLLRPADDTYESYSTIGVDFKIRTELDGKTIKQ 60  
 QY 64 I-----ESFNNVKOMLQEIIDRYASENVKLLVG 91  
 DB 61 IWDVAGQERFRITSSYRGANGICVYDVTDWDSFNNVKOMLQEIIDRYATEGVNKLIVG 120  
 QY 92 NKKDILTKKVVDTTAKERPADSLGIPLETSAKNAATNVEOSEFMTMAEIKRKGAGTAG 151  
 DB 121 NKKDILTKKVVDTTAKERPADSLGIPLETSAKNAATNVEOSEFMTMAEIKRKGAGTAG 180  
 QY 152 GAERKSNVKTISTPVKQSGGGCC 173  
 DB 181 NTKASVNVSPGHGVSNNSGGCC 203  
 RESULT 7  
 YPTL\_CHLRE  
 ID YPTL\_CHLRE STANDARD: PRT; 203 AA.  
 AC 039571;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE GTP-binding protein YPT1.  
 GN YPT1.  
 OS Chlamydomonas reinhardtii.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Chlamydomonadales; Chlamydomonadales; Chlamydomonadales;  
 OC Chlamydomonadales; Chlamydomonadales;  
 OX NCBI\_TaxID=3055;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV45;  
 RX MEDLINE=95309723; PubMed=7789809;  
 RA Dietmaier W., Fabry S., Huber H., Schmitt R.;

FT "Analysis of a family of ypt genes and their products from  
 FT Chlamydomonas reinhardtii";  
 RL Gene 158:41-50(1995).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U13168; AAA82727.1; -  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfrrng.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFAms: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Protein transport;  
 KW Multigene family.  
 FT NP-BIND 15 22 GTP (BY SIMILARITY).  
 FT NP-BIND 63 67 GTP (BY SIMILARITY).  
 FT NP-BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 202 202 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 203 203 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 203 AA; 22599 MW; DD2420EF2F35ELD CRC64;  
 Query Match 73.0%; Score 646.5; DB 1; Length 203;  
 Best Local Similarity 65.7%; Pred. No. 5.3e-47;  
 Matches 134; Conservative 9; Mismatches 26; Indels 35; Gaps 3;  
 QY 4 MNPEVDYLFKLLIGDSGVKSCLLLRPADDTYESYSTIGVDFKIRTELDGKTIKQ 63  
 DB 1 MNPEVDYLFKLLIGDSGVKSCLLLRPADDTYESYSTIGVDFKIRTELDGKTIKQ 60  
 QY 64 I-----ESFNNVKOMLQEIIDRYASENVKLLVG 91  
 DB 61 IWDVAGQERFRITSSYRGANGICVYDVTDWDSFNNVKOMLQEIIDRYASENVKLLVG 120  
 QY 92 NKKDILTKKVVDTTAKERPADSLGIPLETSAKNAATNVEOSEFMTMAEIKRKGAGTAG 149  
 DB 121 NKKDILTKKVVDTTAKERPADSLGIPLETSAKNAATNVEOSEFMTMAEIKRKGAGTAG 180  
 QY 150 AGAERKSNVKTISTPVKQSGGGCC 173  
 DB 181 KAGGPVVRPO-EGKPIKSSSSCC 203  
 RESULT 8  
 YPTL\_SCHPO  
 ID YPTL\_SCHPO STANDARD: PRT; 203 AA.  
 AC P11620;  
 DT 01-OCT-1988 (Rel. 12, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ras-related protein ypt1.  
 GN YPT1 OR SPAC1703.10.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetes; Schizosaccharomycetes;  
 OC Schizosaccharomycetes; Schizosaccharomycetes;  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90228338; PubMed=2328721;  
 RX Miyake S., Tamamoto M.;

RT "Identification of ras-related, YPT family genes in  
 RT Schizosaccharomyces pombe.";  
 RL EMBL J. 9:1417-1422(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-69236485; PubMed-2740223;  
 RA Fawell E., Hook S., Armstrong J.;  
 RT "Nucleotide sequence of a gene encoding a YPT1-related protein from  
 RT Schizosaccharomyces pombe.";  
 RL Nucleic Acids Res. 17:4373-4373(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-972;  
 RX MEDLINE-21848401; PubMed-11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Scouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Shetton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voicakeart G., Aert R., Robben J., Grymoprez B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Filiz C., Holzer E., Moestl D., Hübner H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadiou E., Dreano S., Gloux S., Leclaire V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hunt S.M.,  
 RA Lucas M., Rocher M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shipakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).  
 RN [4]  
 RP ISOPRENOLD.  
 RX MEDLINE-92283842; PubMed-1597466;  
 RA Newman C.M., Giannakouris T., Hancock J.F., Fawell E.H., Armstrong J.,  
 RA Magee A.I.;  
 RT "Post-translational processing of Schizosaccharomyces pombe YPT  
 RT proteins.";  
 RL J. Biol. Chem. 267:11329-11336(1992).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC -----  
 DR EMBL: X52099; CAA36319.1; -;  
 DR EMBL: X15082; CAA3192.1; -;  
 DR EMBL: AL136536; CAB66454.1; -;  
 DR PIR: S04590; S04590.  
 DR PIR: S10025; S10025.  
 DR HSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfams: TIGR00231; small\_gtp; 1.

KW GTP-binding; Lipoprotein; Prenylation; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 202 202 GERANYL-GERANYL.  
 FT LIPID 203 203 GERANYL-GERANYL.  
 FT LIPID 203 203 MNEP -> NMANINR (IN REF. 2).  
 FT CONFLICT 1 4  
 FT CONFLICT 203 203  
 FT CONFLICT 22816 22816 MW; 59B3524F42B83094 CRC64;  
 SQ SEQUENCE 203 AA; 22816 MW; 59B3524F42B83094 CRC64;  
 Query Match 73.08; Score 646.5; DB 1; Length 203;  
 Best Local Similarity 64.38; Pred. No. 5.3e-47;  
 Matches 131; Conservative 16; Mismatches 23; Indels 33; Gaps 2;  
 QY 4 MNPEYDLFLKLLIGDGVGKSCLLRFADDTYESYISTGVDFKRTLELDTKTIKQ 63  
 DB 1 MNPEYDLFLKLLIGDGVGKSCLLRFADDTYESYISTGVDFKRTLELDTKTIKQ 60  
 QY 64 I-----ESFNWKQMLQELDRYASENVKLLVG 91  
 DB 61 IMDTAGGERFRTTSSYRGAGIIVYDTPDQSFNNVKQMLQELDRYAVEGVNRLVG 120  
 QY 92 NKCLTLTKKYVDYTTAAEPADSLGIFPLETSAKNATVEGSEFMMAEIKRMGPATAG 151  
 DB 121 NKSDMVKKVYEVSAEAFADSLNIPPLETSAKDSTVEQAFETMSHOKERMGNTFAS 180  
 QY 152 GAERKNYKI-OSTPVKOSGGCC 173  
 DB 181 SNAKSSVKVGGGTNVSSSSMCC 203  
 RESULT 9  
 YPT1 VOLCA STANDARD; PRT; 203 AA.  
 ID YPT1 VOLCA  
 AC P31584;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE GTP-binding protein ypt1.  
 GN YPT1.  
 OS Volvox carterl.  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Volvocaceae; Volvox.  
 OC NCBI\_TaxID=3067;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-f. Nagariensis / HK10;  
 RX MEDLINE-92380499; PubMed-1511889;  
 RA Fabry S., Nass N., Huber H., Palme K., Jaenicke L., Schmitt R.;  
 RT "The ypt1 gene encodes a small G-protein in the green alga Volvox  
 RT carterl: gene structure and properties of the gene product.";  
 RL Gene 118:153-162(1992).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).  
 CC -----  
 DR EMBL: M93438; AAA34255.1; -;  
 DR PIR: JCI247; JCI247.  
 DR HSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRMG.  
 DR SMART: SM00175; RAB; 1.



RA Matsui M.;  
 RT "Isolation and analysis of cDNAs encoding small GTP-binding proteins  
 of Arabidopsis thaliana.";  
 RL Gene 108:259-264(1991).  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U89959; AAC24370.1; -  
 DR EMBL: D01027; BAA00832.1; -  
 DR PIR: P50279; P50279.  
 DR HSSP: P05713; 3RAB.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfams: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Multigene family.  
 FT NP\_BIND 70 77 GTP (BY SIMILARITY).  
 FT NP\_BIND 118 122 GTP (BY SIMILARITY).  
 FT NP\_BIND 176 179 GTP (BY SIMILARITY).  
 FT DOMAIN 92 100 EFFECTOR REGION (BY SIMILARITY).  
 FT LIPID 235 255 GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 256 256 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 258 AA; 29169 MW; 34D278A0C504638F CRC64;  
 Query Match 69.9%; Score 619.5; DB 1; Length 258;  
 Best Local Similarity 61.6%; Pred. No. 1.2e+44;  
 Matches 15; Conservative 15; Mismatches 30; Indels 33; Gaps 2;  
 QY 3 SNAPEDYLFKLLIGDSGVGKSCLLRFADDTYESYSTIGVDFKRTITELDKTKIKL 62  
 Db 55 TNNPEVDYLFKLLIGDSGVGKSCLLRFSDSYVESYITIGVDFKRTVEQDGKTIKL 114  
 QY 63 QI-----ESFNNVKQWLQETIDRYASENVNKLIV 90  
 Db 115 QIMDTAGQERFRTTSSYYRGAGHIIIVYDTDESFNNVKQWLSEIDRYASDNVNLIV 174  
 QY 91 GNKCDLTTKKVVDYTTAKERADSLGIPLETSAKNATNVQSFMTMAELTKRMGPATG 150  
 Db 175 GNSDLETKNAIPYETAKAFADDEIGIPMETSAKDATNVQAFMAASIKERNASOPAG 234  
 QY 151 GAESKNVKIOSTPVKOSGGCC 173  
 Db 235 NNARPTVQIRGQPVNOK-NGCC 256  
 RESULT 12  
 YPT2\_MAIZE  
 ID YPT2\_MAIZE STANDARD; PRT; 203 AA.  
 AC 005737;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE GTP-binding protein YPTM2.  
 GN YPTM2.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC Clade;  
 OC Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Coleoptile;  
 RX MEDLINE=92115746; PubMed=1731354;

RA Palme K., Diefenthal T., Vingron M., Sander C., Schell J.;  
 RT "Molecular cloning and structural analysis of genes from Zea mays  
 (L.) coding for members of the ras-related ypt gene family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:787-791(1992).  
 CC -1- FUNCTION: PROTEIN TRANSPORT. PROBABLY INVOLVED IN VESICULAR  
 CC TRAFFIC (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: ITS EXPRESSION IS WEAK IN STEMS, HIGHER IN  
 CC ROOTS, LEAVES AND COLEOPTILES, BUT HIGHEST IN FLOWERS.  
 CC -1- PTM: THE PALMITIC ACID IS REQUIRED FOR MEMBRANE ATTACHMENT AND  
 CC BIOLOGICAL FUNCTION.  
 CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X63278; CAA44919.1; -  
 DR PIR: B38202; B38202.  
 DR HSSP: P05713; 3RAB.  
 DR MaizeDB: 78605; -  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnsfmg.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; ras; 1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRfams: TIGR00231; small\_GTP; 1.  
 KW GTP-binding; Lipoprotein; Prenylation; Palmitate; Protein transport.  
 FT NP\_BIND 15 22 GTP (BY SIMILARITY).  
 FT NP\_BIND 63 67 GTP (BY SIMILARITY).  
 FT NP\_BIND 121 124 GTP (BY SIMILARITY).  
 FT DOMAIN 37 45 EFFECTOR REGION (PROBABLE).  
 FT LIPID 200 200 PALMITATE (BY SIMILARITY).  
 FT LIPID 201 201 GERANYL-GERANYL (BY SIMILARITY).  
 SQ SEQUENCE 203 AA; 22475 MW; E241326E7ACD1B8A CRC64;  
 Query Match 69.8%; Score 618.5; DB 1; Length 203;  
 Best Local Similarity 63.4%; Pred. No. 1.1e+44;  
 Matches 128; Conservative 11; Mismatches 30; Indels 33; Gaps 2;  
 QY 4 MNPEYDYLKLLIGDSGVGKSCLLRFADDTYESYSTIGVDFKRTITELDKTKIKL 63  
 Db 1 MNPEYDYLKLLIGDSGVGKSCLLRFADDTYLDSTISTIGVDFKRTVEQDGKTIKL 60  
 QY 64 I-----ESFNNVKQWLQETIDRYASENVNKLIV 91  
 Db 61 IMDTAGQERFRTTSSYYRGAGHIIIVYDTDESFNNVKQWLSEIDRYASDNVNLIV 120  
 QY 92 NKCDLTTKKVVDYTTAKERADSLGIPLETSAKNATNVQSFMTMAELTKRMGPATG 151  
 Db 121 NKSDLTANKVVAETETAFAFDEMGIPMETSAKNATNVQAFMAASIKRNASOPAAA 180  
 QY 152 GAESKNVKIOSTPVKOSGGCC 173  
 Db 181 NNARPTVQIRGQPVNOK-TSCC 201  
 RESULT 13  
 RIC1\_ORYZA  
 ID RIC1\_ORYZA STANDARD; PRT; 202 AA.  
 AC P40392;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ras-related protein RIC1.  
 GN RIC1.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;



```

FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFECTOR REGION (PROBABLE).
FT LIPID 205 205 PALMITATE.
FT LIPID 206 206 GERANYL-GERANYL.
FT MUTAGEN 37 37 Y->F: NO CHANGE.
FT MUTAGEN 39 39 S->A: NO CHANGE.
FT MUTAGEN 40 40 T->S: NO CHANGE.
FT MUTAGEN 41 41 I->M: LETHAL.
FT MUTAGEN 43 43 V->E: NO CHANGE.
FT MUTAGEN 44 44 D->N: TEMPERATURE-SENSITIVE PHENOTYPE.
FT MUTAGEN 121 121 N->I: LETHAL.
FT CONFLICT 171 171 E -> Q (IN REF. 1).
SQ SEQUENCE 206 AA; 23214 MW; F8C704F6BF2D227B CRC64;

Query Match 64.4%; Score 571; DB 1; Length 206;
Best Local Similarity 56.3%; Pred. No. 1e-40;
Matches 116; Conservative 21; Mismatches 33; Indels 36; Gaps 2;

QY 4 MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDPKIRITELDGKTIKQ 63
DB 1 MNSDYDLFKLLIGDSGVGKSCLLRFSDTYINDYISTIGVDPKIRITELDGKTIKQ 60
QY 64 I-----ESFNNVQWLOEIDRYASENVNKLIVG 91
DB 61 IMDTAGEFRRTITSYRGSHGIIVYDVTDGSEFNGVKWMLQEIDRYATSTVLKLLV 120
QY 92 NKCDLTTRKKVYDTTAKAFADSLGIPLETSAKNAATVEQSFMMAEIKRMP-----G 147
DB 121 NKCDLKRVEYDVAKAFADANKMPLETSAIDSTVEAFMLMAQIKESMSOONLINE 180
QY 148 ATAGAERSNVKIQSTPVKOSGGCC 173
DB 181 TTQKEDKGNVNLKGSLTNTGGCC 206

RESULT 15
RBA1_DICDI STANDARD; PRT; 167 AA.
AC P34139;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Ras-related protein RAB1A.
GN RABA OR RAB1A.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-AX3;
RX MEDLINE=94124041; PubMed=8294041;
RA Bush J.M. IV, Franek K., Daniel J.M., Spiegelman G.B., Weeks G.,
RA Cardelli J.A.;
RT "Cloning and characterization of five novel Dictyostelium discoideum
RT rab-related genes.";
RL Gene 136:55-60(1993).
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L21009; AAC37385.1; -
DR HSSP: P05713; 3RAB.
DR DictyDb: DD05055; raba.
DR InterPro: IPR003579; GTPase_Rab.
DR InterPro: IPR001806; Ras_trnsfmg.

```

```

DR InterPro: IPR005225; Small-GTP.
DR Pfam: PF00071; Ras; 1.
DR SMART: SM00175; RAB; 1.
DR TRIGRFAMS: TRIGR00231; small_GTP; 1.
KW GTP-binding.
FT NP_BIND 15 22 GTP (BY SIMILARITY).
FT NP_BIND 63 67 GTP (BY SIMILARITY).
FT NP_BIND 121 124 GTP (BY SIMILARITY).
FT DOMAIN 37 45 EFECTOR REGION (BY SIMILARITY).
SQ SEQUENCE 167 AA; 18943 MW; 72480B04D943BD8D CRC64;

Query Match 61.2%; Score 542; DB 1; Length 167;
Best Local Similarity 67.7%; Pred. No. 2e-38;
Matches 111; Conservative 10; Mismatches 11; Indels 32; Gaps 1;

QY 4 MNPEYDYLFKLLIGDSGVGKSCLLRFADDTYESYISTIGVDPKIRITELDGKTIKQ 63
DB 1 MNPDYHLFKLLIGDSGVGKSCLLRFADDTSESISTIGVDPKIRITELDGKTIKQ 60
QY 64 I-----ESFNNVQWLOEIDRYASENVNKLIVG 91
DB 61 IMDTAGEFRRTITSYRGSHGIIVYDVTDKLFENVRQWMLQEIDRFACENVNKLIVG 120
QY 92 NKCDLTTRKKVYDTTAKAFADSLGIPLETSAKNAATVEQSFMT 135
DB 121 NKSDLVAKKVDENTAKAFADSLQIPLETSAKOSTVEQAFMT 164

```

Search completed: November 17, 2002, 13:05:10  
Job time : 26 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: November 17, 2002, 12:50:47 ; Search time 81 Seconds

(without alignments)  
440.076 Million cell updates/sec

Title: US-09-820-003a-2

Perfect score: 886  
Sequence: 1 MSXMPDYFLKLLIGDS.....EKSNNKIOSTFVKGSGGCC 173Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mmc:\*  
8: SP-organelle:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-virus:\*  
13: SP-vertebrate:\*  
14: SP-unclassified:\*  
15: SP-virus:\*  
16: SP-bacteria:\*  
17: SP-archaea:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	766.5	86.5	201	4 Q9H0U4	Q9H0U4 homo sapien
2	759.5	85.7	201	11 Q9NDI1	Q9NDI1 mus musculu
3	700	79.0	141	4 Q96N61	Q96N61 homo sapien
4	683	77.1	205	5 O18332	O18332 drosophila
5	671	75.7	202	5 O16143	O16143 bombyx mori
6	660	74.5	205	5 Q9U406	Q9U406 caenorhabdi
7	655.5	74.0	201	3 Q9HET3	Q9HET3 aspergillus
8	653.5	73.8	201	3 Q9P8J7	Q9P8J7 aspergillus
9	644	72.7	202	10 Q40203	Q40203 lotus japon
10	642	72.5	202	3 Q9SEH3	Q9SEH3 arabidopsis
11	640	72.2	202	3 Q9HDT5	Q9HDT5 trichoderma
12	636.5	71.8	203	10 Q8M154	Q8M154 pisum sativ
13	634.5	71.6	202	10 Q8W4S8	Q8W4S8 arabidopsis
14	633	71.4	202	10 Q9FPJ4	Q9FPJ4 arabidopsis
15	632.5	71.4	203	10 Q24112	Q24112 nicotiana p
16	632	71.3	202	10 Q9SXT5	Q9SXT5 clever arlet

17	628.5	70.9	203	10 Q40204	Q40204 lotus japon
18	627.5	70.8	203	10 Q40569	Q40569 nicotiana t
19	624.5	70.5	203	10 Q40205	Q40205 lotus japon
20	624	70.4	129	4 Q9Y3T2	Q9Y3T2 homo sapien
21	623	70.3	202	10 Q39845	Q39845 glycine max
22	622.5	70.3	203	10 Q9W7P5	Q9W7P5 papilion an
23	622.5	70.3	203	10 Q41338	Q41338 lycopersico
24	622.5	70.3	203	10 Q41340	Q41340 lycopersico
25	622.5	70.1	203	10 Q419E2	Q419E2 oryza sativ
26	621.5	69.8	203	10 Q90155	Q90155 pisum sativ
27	618.5	69.8	203	10 Q90157	Q90157 arabidopsis
28	616	69.5	202	10 Q08153	Q08153 pisum sativ
29	612	69.1	221	10 Q23594	Q23594 oryza sativ
30	608	68.6	203	10 Q8R063	Q8R063 oryza sativ
31	598	67.5	218	10 Q39861	Q39861 glycine max
32	596.5	67.3	207	3 Q96W97	Q96W97 candida alb
33	590.5	66.6	202	10 Q9ZRH6	Q9ZRH6 petunia hyb
34	589	66.5	201	10 Q08152	Q08152 pisum sativ
35	589	66.5	201	10 Q9L6G3	Q9L6G3 oryza sativ
36	587	66.3	201	10 Q43462	Q43462 glycine max
37	583.5	65.9	202	10 Q41339	Q41339 lycopersico
38	581	65.6	201	5 Q25324	Q25324 lotus japon
39	580	65.5	200	10 Q39274	Q39274 brassica ca
40	574	64.8	206	10 Q9ZRE2	Q9ZRE2 arabidopsis
41	569.5	64.3	205	10 Q9ZRE2	Q9ZRE2 lotus japon
42	569	64.2	194	10 Q40202	Q40202 lotus japon
43	563	63.5	206	3 Q9HFM5	Q9HFM5 knuyveromyc
44	503	56.8	200	5 Q9BNH1	Q9BNH1 plasmodium
45	484	54.6	205	5 Q9BLF3	Q9BLF3 entamoeba h

## ALIGNMENTS

## RESULT 1

Q9H0U4 PRELIMINARY: PRT: 201 AA.

AC Q9H0U4: 01-MAR-2001 (TREMUREL, 16, Created)  
DT 01-MAR-2001 (TREMUREL, 16, Last sequence update)  
DT 01-JUN-2002 (TREMUREL, 21, Last annotation update)  
DE Hypothetical 22.2 kDa protein.  
CN DKFZ5641172  
OS Homo sapiens (human)  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=21134917; PubMed=11230166;  
RA Wiesmann S., Weil B., Wellenreuther R., Gassenhuber J., Glasel S.,  
RA Ansoerger W., Boecker M., Bloeker H., Bauersachs S., Blum H.,  
RA Lauder J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
RA Mewes H.W., Olterswaelder B., Obermaier B., Tampe J., Heuener D.,  
RA Wambutt R., Korn B., Klein M., Poustka A.;  
RT Towards a Catalog of Human Genes and Proteins: Sequencing and  
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs.\*;  
RL Genome Res. 11:422-435(2001).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAS SUBFAMILY.  
DR EMBL: AL136635; CAB6570.1; -;  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsfrmg.  
DR InterPro: IPR002078; S1954\_interact.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; ras; 1.  
DR PRINTS: PRO0449; RASTRNSFRMG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRPFAM: TIGR0231; small\_gtp; 1.  
DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
KW GTP-binding; Hypothetical protein; Lipoprotein.  
SQ SEQUENCE 201 AA: 22171 MW: 9812FADAC34B2BE CRC64;



121 TSANKATNVEQSPMTAAEIKRMKPGATAGCAEKSNNKIOSIPVKGSGGCC 173  
 89 TSANKATNVEQSPMTAAEIKRMKPGATAGCAEKSNNKIOSIPVKGSGGCC 141  
 RESULT 4  
 018332 PRELIMINARY: PRT: 205 AA.  
 ID 018332  
 AC 018332  
 DT 01-JAN-1998 (TRENBLREL 05, Created)  
 DT 01-JAN-1998 (TRENBLREL 05, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL 21, Last annotation update)  
 DE RAB1 protein  
 GN RAB1 OR CG3320  
 OS Drosophila melanogaster (Fruit fly)  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidea; Drosophilidae; Drosophila  
 NC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY  
 RX MEDLINE=20196006; PubMed-10731132;  
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,  
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borikova D., Botchan M.R., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Burks J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen T.,  
 RA Splet E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtikas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhou L.,  
 RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhou S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-OREGON R: TISUE-HEAD.  
 RX MEDLINE=97228579; PubMed-9074639;  
 RA Satoh A.K., Tokunaga F., Ozaki K.,  
 RT "Rab proteins of Drosophila melanogaster: novel members of the Rab-  
 protein family."  
 RL FEBS Lett. 404:65-69(1997).  
 CC 1 SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL: AB03734; AF55873.1; -.  
 DR EMBL: D84312; BAA21705.1; -.

DR HSSP: P05713; 3RAB.  
 DR FlyBase: FBgn0016700; Rab1.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnstrng.  
 DR InterPro: IPR002078; S1G54\_interact.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFS: TIGR0231; sma11\_GTP; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 DR GTP-binding, Lipoprotein.  
 KW GTP-binding, 205 AA; 22762 MW; 9CE2691205FAE343 CRC64;  
 SQ SEQUENCE  
 Query Match 77.1%; Score 683; DB 5; Length 205;  
 Best Local Similarity 69.4%; Pred. No. 2, 2e-47;  
 Matches 143; Conservative 12; Mismatches 17; Indels 34; Gaps 3;  
 1 MSSMPEVDYLFKLLIGDSGVKSCLLRFADDTYESYSTTGVDKIRTEIDGKTI 60  
 1 MSSVPEVDYLFKLLIGDSGVKSCLLRFADDTYESYSTTGVDKIRTEIDGKTI 60  
 Db 1 MSSVPEVDYLFKLLIGDSGVKSCLLRFADDTYESYSTTGVDKIRTEIDGKTI 60  
 61 K101-----ESFNNVKOMLOEIDRYASENVKL 88  
 Db 61 K101DVTAGQERFTTSSYRGAGLIYYDCTDSEFNKKOMLEETERRACENVKL 120  
 89 LVGNKCDLTTKKVVDYTTAKEPADSLGIPLETSAKNATNVEQSPMTAAEIKRMKPGCA 148  
 Db 121 LVGNKSDLTTRKKVVDHTTAETAAQDGIPLFETSAKSATNVEQAFMTAAEIKRMVGPSS 180  
 149 TAGAEKSNVKI -GSTPVKQSGGCC 173  
 Db 181 SATD -NASKVKIDQGRVENTKSGCC 205  
 RESULT 5  
 016143  
 ID 016143 PRELIMINARY: PRT: 202 AA.  
 AC 016143  
 DT 01-JAN-1998 (TRENBLREL 05, Created)  
 DT 01-JAN-1998 (TRENBLREL 05, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL 21, Last annotation update)  
 DE Small GTP-binding protein.  
 GN RAB8  
 OS Bombyx mori (Silk moth)  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
 OC Bombycoidea; Bombycidae; Bombyx.  
 NC NCBI\_TaxID=7091.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Uno T., Nakajima A., Ueno M., Aizono Y.,  
 RT "Isolation of cDNA encoding small GTP-binding protein from Bombyx  
 mori.";  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 CC 1-1 SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
 DR EMBL: AF013572; BAB67169.1; -.  
 DR HSSP: P05713; RAB8; 1.  
 DR InterPro: IPR003579; GTPase\_Rab.  
 DR InterPro: IPR001806; Ras\_trnstrng.  
 DR InterPro: IPR002078; S1G54\_interact.  
 DR InterPro: IPR005225; Small\_GTP.  
 DR Pfam: PF00071; Ras\_1.  
 DR PRINTS: PR00449; RASTRNSFRNG.  
 DR SMART: SM00175; RAB; 1.  
 DR TIGRFS: TIGR0231; sma11\_GTP; 1.  
 DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
 DR GTP-binding, Lipoprotein.  
 KW GTP-binding, 202 AA; 22364 MW; 0338E95D63CDFE4 CRC64;  
 SQ SEQUENCE  
 Query Match 75.7%; Score 671; DB 5; Length 202;  
 Best Local Similarity 66.5%; Pred. No. 2e-46;  
 Matches 135; Conservative 18; Mismatches 16; Indels 34; Gaps 3;

QY	4	MNEPYDLFLLILIGSGVKSCLLRADDPYTESYSTGVDFKRTTELDGKITKO	63
Dd	1	MNPEYDULFKLLILIGSGVKSCLLRADDPYTESYSTGVDFKRTTELDGKITKO	60
QY	64	I-----ESFNVKQMIOEIDRYASENVKNLVLG	91
Dd	61	IMPTAGGERFRITSSYYRGAGIIIVYDCDDODSFNSVKMKWMEEDIEIRYACDNVNKLVLG	12
QY	92	NKCQLTKKKVVDYDTTKEEFADSLGIPILETSAKNAANTNESPFMTMAELIKRMGPATNG	15
Dd	121	NKCQLTKKKVVDYDTTAKOYAEOGLIPILETSAKNSTNEAFMTMAELIKARVPSPST-G	17
QY	152	GAEKSNYKI-QSTPVKQSGGGCC	173
Dd	180	AAPAGHYKIDQGQPIDTGKSSCC	202
RESULT	6		
Q9UOA6	09UOA6	PRELIMINARY;	PRT; 205 AA.
AC	09UOA6:		
DT	01-MAY-2000 (TREMBLrel_13, Created)		
DT	01-MAY-2000 (TREMBLrel_13, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel_21, Last annotation update)		
DE	Hypothetical 22.5 kDa protein.		
GN	C39F7.4.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea;		
OC	Rhabdilitidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_Taxid=6239;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RC	SCHNEE P., MAGGI U., DUBBELDE C.;		
RC	"The sequence of C. elegans cosmid C39F7."		
RA	The sequence of C. elegans cosmid C39F7."		
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RP	STRAIN-BRISTOL N2;		
RC	SPRAIN-BRISTOL N2;		
RA	Waterston R.;		
RA	"Direct Submission."		
RL	Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.		
CC	-I- SIMILIARTY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.		
DR	EMBL: AF010310. AAC69218.1; -		
DR	HSSP: P01112; IPTL.		
DR	InterPro: IPR003579; GTPase_Rab.		
DR	InterPro: IPR001806; Ras_trnstrng.		
DR	InterPro: IPR002078; Slg54_interact.		
DR	InterPro: IPR005225; Sma11_GTP.		
DR	Pfam: PF000071; ras: 1.		
DR	PRINTS: PR00449; RASTRNSFRMNG.		
DR	SMART: SM00175; RAB; 1.		
DR	TIGRFAMs: TIGR00231; sma11_gtp_1.		
DR	PROSITE: PS00675; SIAMA54_INTERACT_1; UNKNOWN_1.		
KW	GTP-binding; Hypothetical protein; Lipoprotein.		
SO	SEQUENCE 205 AA; 22545 MW; 5221920FP0584203 CRC64;		
Query Match	74.5%; Score 660; DB 5; Length 205;		
Best Local Similarity	66.2%; Pred. NO. 1.3e-45;		
Matches 137; Conservative 16; Mismatches 18; Indels 36; Gaps			
QY	1	MSSNNPEYDLFKLLILIGSGVKSCLLRADDPYTESYSTGVDFKRTTELDGKIT	66
Dd	1	MAAANPEYDLFKLLILIGSGVKSCLLRADDPYTESYSTGVDFKRTTELDGKIT	60

QY	61	KUOI	-----	ESFNNNKQWLOEIDRYASENVNKL	88
Db	61	KUOIDYDFAQGEFRITTSYYNGAIGIIVYDITDQEFITNNVKNQWLOEIDRACRACENVKL	120		
QY	89	LVGNKCDLTKRKVVYDTTAKAFADSLGIPFLETSKAKNATNVBQSFMTMAELTKRMGPCA	148		
Db	121	LVGNKCDLTKRAKAVETQAADYAGOLGIPFLETSKAKSNVNBQAFELTMAELTKSRMGPAVO	180		
QY	149	TAGGAEKSNVVKIQ-STPYK-QSGGGCC	173		
Db	181	GAGGA--PCVRRITGSPQVODKKSGGCC	205		
RESULT 7					
Q9HER3	Q9HER3	PRELIMINARY:	PRT:	201 AA.	
AC	Q9HER3:				
DT	01-MAR-2001 (TREMBLrel. 16, Created)				
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Secretion related GTPase (Srgb).				
GN	Scgb.				
OS	Aspergillus niger.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
CC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.				
OX	NCBI_TaxID=5061;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=N402;				
RA	Punt P.J., Selboth B., Weenink X.O., van Zeijl C.M., Lenders M.,				
RA	Konetschny C., Ram A.F., Montijn R., Kubicek C.P.,				
RA	Van den Hondel C.A.;				
RT	"Identification and characterisation of a family of secretion related				
RT	small GTPase encoding genes from the filamentous fungus Aspergillus				
RT	niger: a putative Sec4 homologue is not essential for growth.";				
RL	Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.				
CC	-1 SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.				
DR	EMBL: AJ278659; CAC17833.1; -				
DR	HSSP: P01112; 1p1j				
DR	InterPro: IPR003579; GTPase_Rab.				
DR	InterPro: IPR001806; Ras_trnstrimng.				
DR	InterPro: IPR02078; Slg54_interact.				
DR	InterPro: IPR005225; Small_GTP.				
DR	Pfam: PF00071; ras; 1.				
DR	PRINTS: PR00449; RASTRNSFRMNG.				
DR	SMART: SM00175; RAB; 1.				
DR	TIGRFAMs: TIGR00231; small_gtp_1.				
DR	PROSITE: PS00675; SIGMAS54_INTERACT_1; UNKNOWN_1.				
KW	GTP-binding; Lipoprotein.				
SO	SEQUENCE 201 AA; 22303 MW; 80C1A5B0BBDP378 CRC64;				
Query Match 74.0%; Score 655.5; DB 3; Length 201;					
Best Local Similarity 67.2%; Pred. No. 3,4e-45;					
Matches 137; Conservative 12; Mismatches 18; Indels 37; Gaps 3					
QY	4	NMPEDYLEKLLILIGDSGVGKSCLLRFADDDYTESYISTIGVDFKIRITELDGKTIKIQ	63		
Db	1	NMPEDYLEKLLILIGDSGVGKSCLLRFADDDYTESYISTIGVDFKIRITELDGKTIKIQ	60		
QY	64	I-----ESENNNKQWLOEIDRYASENVNKL	91		
Db	61	INDVGAQGEFRITTSYYNGAIGICVYDITDQEFITNNVKNQWLOEIDRATSGVNNKL	120		
QY	92	NKCDLTKRKVVYDTTAKAFADSLGIPFLETSKAKNATNVBQSFMTMAELTKRMGPATG	151		
Db	121	NKSDMEDKKVVEYTAKEFADSLGIPFLETSKAKSNVNBQAFELTMAELTKSRMG--TAF	177		
QY	152	GAEKSNVVKIQSTPYKQSG--GGCC	173		
Db	178	VNNKPTVOVGQGVQSGSGAGGCC	201		

RESULT 8  
Q9PB77 PRELIMINARY: PRT: 201 AA.  
ID Q9PB77: PRELIMINARY: PRT: 201 AA.  
AC Q9PB77: PRELIMINARY: PRT: 201 AA.  
BT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE YPA.  
GN YPA.  
OS Aspergillus awamori.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI\_TaxID=103351;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Saloheimo M., Wang H., Valkonen M., Vasara T., Huuskonen A., Ward M.,  
RA Penttila M.,  
RA The secretory genes ypt1/ypt4 and nsf1/nsf4 from the filamentous  
RT fungal Trichoderma reesei and Aspergillus niger: evidence for global  
RT transcriptional regulation of the secretory pathway in Trichoderma  
RT reesei.  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC EMBL: AF244545; AAF63333.1; -.  
CC HSSP: P01112; 1pL1.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_crasfmg.  
DR InterPro: IPR002078; Sig54\_Interact.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; ras; 1.  
DR PRINTS: PR00449; RASTRNSFRMNG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRFS: TIGR00231; small\_gtp; 1.  
DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
KM GTP-binding; Lipoprotein.  
SQ SEQUENCE 201 AA: 22335 MW: 9D710F1ABEBE279 CRC64;  
Query Match 73.8%; Score 653.5; DB 3; Length 201;  
Best Local Similarity 67.2%; Pred. No. 56-45; Mismatches 18; Indels 37; Gaps 3;  
Matches 137; Conservative 12; Mismatches 18; Indels 37; Gaps 3;  
DB 1 MNEVDYLFKLLIGDSGVKSCLLRPADDTYESTISTGVDFKIRTELDGKTIKQ 63  
QY 4 MNEVDYLFKLLIGDSGVKSCLLRPADDTYESTISTGVDFKIRTELDGKTIKQ 63  
BT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE RAS-related small GTP-binding protein (RAS-related small GTP-binding  
DE protein RAB1C) (A0461750/DLA800C).  
DE RAB1C OR A0461750.  
GN Arabidopsis thaliana (Mouse-ear cress).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Straub C.V., Coudreau J.,  
RA Bischoff F., Goode M., Palme J.;  
RT "Arabidopsis involved in ER-Golgi traffic in meristematic cells in  
RT Arabidopsis." (Feb-1999) to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk K., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,

RP SEQUENCE FROM N.A.  
RC TISSUE-ROOT NODULES;  
RX MEDLINE-97231679; PubMed-9076991;  
RA Borg S., Brandstrup B., Jensen T.J., Poulsen C.;  
RA "Identification of new protein species among 33 different small GTP-  
RT binding proteins encoded by cDNAs from Lotus japonicus, and expression  
RT of corresponding mRNAs in developing root nodules.";  
RT plant J. 11:237-250(1997).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
CC EMBL: Z73932; CAA98160.1; -.  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_crasfmg.  
DR InterPro: IPR002078; Sig54\_Interact.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; ras; 1.  
DR PRINTS: PR00449; RASTRNSFRMNG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRFS: TIGR00231; small\_gtp; 1.  
DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
KM GTP-binding; Lipoprotein.  
SQ SEQUENCE 202 AA: 22633 MW: 91633BEPAD07662F CRC64;  
Query Match 72.7%; Score 644; DB 10; Length 202;  
Best Local Similarity 65.8%; Pred. No. 2.9e-44; Mismatches 27; Indels 34; Gaps 3;  
Matches 133; Conservative 8; Mismatches 27; Indels 34; Gaps 3;  
DB 1 MNEVDYLFKLLIGDSGVKSCLLRPADDTYESTISTGVDFKIRTELDGKTIKQ 63  
QY 4 MNEVDYLFKLLIGDSGVKSCLLRPADDTYESTISTGVDFKIRTELDGKTIKQ 63  
BT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE RAS-related small GTP-binding protein (RAS-related small GTP-binding  
DE protein RAB1C) (A0461750/DLA800C).  
DE RAB1C OR A0461750.  
GN Arabidopsis thaliana (Mouse-ear cress).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Straub C.V., Coudreau J.,  
RA Bischoff F., Goode M., Palme J.;  
RT "Arabidopsis involved in ER-Golgi traffic in meristematic cells in  
RT Arabidopsis." (Feb-1999) to the EMBL/GenBank/DBJ databases.  
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk K., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,

RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis cDNA clones";  
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL: AF127134; CAA722133.1; -  
DR EMBL: A161546; CAB78756.1; -  
DR EMBL: AY052204; AAK97673.1; -  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsmg.  
DR InterPro: IPR002078; Sig54\_interact.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; Ras; 1.  
DR PRINTS: PR00449; RASTRNSFRNG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRFAMs: TIGR00231; small\_gtp; 1.  
DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
KW GTP-binding; Lipoprotein.  
KW VARIANT 70  
SQ SEQUENCE 202 AA; 22318 MW; 954FA24C3110FC12 CRC64;

Query Match 72.5%; Score 642; DB 10; Length 202;  
Best Local Similarity 65.8%; Pred. No. 4.2e-44;  
Matches 133; Conservative 11; Mismatches 24; Indels 34; Gaps 3;

QY 4 MNPEYDYLFKLLIGDSGVGSKCLLRPADTYTESYSTIGVDFKRTIELDKTKLQ 63  
DB 1 MNPEYDYLFKLLIGDSGVGSKCLLRPADSYLDSYSTIGVDFKRTIELDKTKLQ 60  
QY 64 I-----ESFNVKQWLOEIDRYASENVKLLVG 91  
DB 61 IWDTAGOERFRTITSSYRGAGHICVYDVTDMDSFNVKQWLOEIDRYASENVKLLVG 120  
QY 92 NKCDLTKKVVDTYTAKEFADSLGIPLETSAKNATNEQSFMTMAEIKRKGCGATAG 151  
DB 121 NKCDLTSKVVSTETAKFADSLGIPLETSAKNATNEQAFMTAIAIKTRMA-SQPAG 179  
QY 152 GAESKNVAKIOSTPVKOSGGGCC 173  
DB 180 GSKPPTVQIRGQPVNQ-SGCC 200

RESULT 11  
Q9HDT5 PRELIMINARY; PRT; 202 AA.  
AC Q9HDT5;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Small GTP-binding protein YPT1.  
OS Trichoderma reesei (Hypocrea jecorina).  
CC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
CC Hypocreales; Hypocreaceae; Hypocrea.  
OX NCBI\_TaxID=51453;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RTCC-30;  
RA Salohelmo M.L.A., Mann H., Valkonen M., Vasara T., Huuskonen A.,  
RA Rikonen M., Pakula T., Ward M., Penttilae M.;  
RT "The secretory genes ypt1/ypt4 and nsf1/nsf4 from the filamentous  
fungi Trichoderma reesei and Aspergillus niger var. awamori: evidence  
for global transcriptional regulation of the secretory pathway in T.  
reesei";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL: AJ277108; CAC17744.1; -

DR HSSP: P01112; 1PLJ.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsmg.  
DR InterPro: IPR002078; Sig54\_interact.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; Ras; 1.  
DR PRINTS: PR00449; RASTRNSFRNG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRFAMs: TIGR00231; small\_gtp; 1.  
DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
KW GTP-binding; Lipoprotein.  
SQ SEQUENCE 202 AA; 22421 MW; A60E10E2AB1B2E7 CRC64;

Query Match 72.2%; Score 640; DB 3; Length 202;  
Best Local Similarity 65.7%; Pred. No. 6.1e-44;  
Matches 134; Conservative 13; Mismatches 21; Indels 36; Gaps 4;

QY 4 MNPEYDYLFKLLIGDSGVGSKCLLRPADTYTESYSTIGVDFKRTIELDKTKLQ 63  
DB 1 MNPEYDYLFKLLIGDSGVGSKCLLRPADTYTESYSTIGVDFKRTIELDKTKLQ 60  
QY 64 I-----ESFNVKQWLOEIDRYASENVKLLVG 91  
DB 61 IWDTAGOERFRTITSSYRGAGHICVYDVTDMDSFNVKQWLOEIDRYASENVKLLVG 120  
QY 92 NKCDLTKKVVDTYTAKEFADSLGIPLETSAKNATNEQSFMTMAEIKRKGCGATAG 151  
DB 121 NKSDMSKVVETVYAKFADSLGIPLETSAKNATNEQAFMTAIAIKTRMA-GTTTA 178  
QY 152 GAESKNVAKI-OSTPV-KOSGGGCC 173  
DB 179 NNTKPSVHVGGGCGVGNSSSSCC 202

RESULT 12  
Q08154 PRELIMINARY; PRT; 203 AA.  
AC Q08154;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE GTP-binding protein.  
OS Pisum sativum (Garden pea).  
CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
CC eucosids; I; Fabales; Fabaceae; Papilionoideae; Viciaceae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LEAF;  
RX MEDLINE=94290824; PubMed=6019783;  
RA Nagano Y., Mural N., Matsuno R., Sasaki Y.;  
RT "Isolation and characterization of cDNAs that encode eleven small GTP-  
binding protein from Pisum sativum";  
RL Plant Cell Physiol. 34:447-455(1993).  
CC -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.  
DR EMBL: D12549; BAA02117.1; -  
DR HSSP: P05713; 3RAB.  
DR InterPro: IPR003579; GTPase\_Rab.  
DR InterPro: IPR001806; Ras\_trnsmg.  
DR InterPro: IPR002078; Sig54\_interact.  
DR InterPro: IPR005225; Small\_GTP.  
DR Pfam: PF00071; Ras; 1.  
DR PRINTS: PR00449; RASTRNSFRNG.  
DR SMART: SM00175; RAB; 1.  
DR TIGRFAMs: TIGR00231; small\_gtp; 1.  
DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.  
KW GTP-binding; Lipoprotein.  
SQ SEQUENCE 203 AA; 22463 MW; 6F2E1007A31EF34 CRC64;

Query Match 71.8%; Score 636.5; DB 10; Length 203;  
Best Local Similarity 64.9%; Pred. No. 1.2e-43;  
Matches 131; Conservative 11; Mismatches 27; Indels 33; Gaps 2;

RESULT 13

DB 152 GAESNVKIQSTPVKQSGGCC 173

DB 181 NRPPTVQIRGQPVNOQSG -GCCC 201

PRELIMINARY: PRT: 202 AA.

AC 08M4S8

DB 01-MAR-2002 (TREMREL. 20, Created)

DB 01-MAR-2002 (TREMREL. 20, Last sequence update)

DB 01-JUN-2002 (TREMREL. 21, Last annotation update)

AT4917530/d14800C.

OS Arabidopsis thaliana (Mouse-ear cress): Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Streptophyta; eudicotyledons; core eudicots; Rosidae; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsids.

NCBI\_TaxID=3702;

SEQUENCE FROM N.A.

Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J., Bower L., Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamaya A., Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yoshizaki K., Davis R.W., Theologis A., Ecker J.R.;

"Arabidopsis ORF clones.";

Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

EMBL: AY060495; AAL31108.1; -

InterPro: IPR003579; GTPase\_Rab.

InterPro: IPR001806; Ras\_transfmg.

InterPro: IPR002078; S1954\_interact.

Pfam: PF00071; ras; 1.

PRINTS: PRO0449; RASTRNSFRMG.

SMART: SM00175; RAB; 1.

PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.

SEQUENCE 202 AA; 22286 MW; 954FAL2C3110FC12 CRC64;

Query Match

Best Local Similarity 71.6%; Score 634.5; DB 10; Length 202;

Best Local Similarity 65.8%; Pred. No. 1.7e-43;

Matches 131; Conservative 11; Mismatches 24; Indels 33; Gaps 2;

DB 4 MNEPDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRIETLDEGKTILQ 63

DB 1 MNEPDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRIETLDEGKTILQ 60

DB 64 I-----ESFNVMKQMLEIDRYASENVNKLIVG 91

DB 61 IMDTAGGERFRTTSSYRGAGHIIYVDTLSEFNVMKQMLEIDRYASENVNKLIVG 120

DB 92 NKCDLTKKVVDTTAKAFADSLGIPLETSAKNAITNVEOSFPMMAEIKRMGCGATAG 151

DB 121 NKCDLTSOKVSTETAKAFADSLGIPLETSAKNAITNVEAFMATAITRMA -SOPAG 179

DB 152 GAESNVKIQSTPVKQSGGCC 170

DB 180 GSKPPTVQIRGQPVNOQSG 198

RESULT 14

DB 09FPI4

DB 09FPI4

PRELIMINARY: PRT: 202 AA.

AC 08FPI4

DB 01-MAR-2001 (TREMREL. 16, Created)

DB 01-MAR-2001 (TREMREL. 16, Last sequence update)

DB 01-JUN-2002 (TREMREL. 21, Last annotation update)

AT5647200 (Putative ras-related small GTP-binding protein).

AT5647200/MOL5.5 OR AT5647200.

OS Arabidopsis thaliana (Mouse-ear cress): Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Streptophyta; eudicotyledons; core eudicots; Rosidae; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eucosids II; Brassicales; Brassicaceae; Arabidopsids.

NCBI\_TaxID=3702;

SEQUENCE FROM N.A.

Shim P., Chao Q., Brooks S., Chen H., Cheuk R., Johnson-Hopson C., Khan S., Kim C.J., Banh J., Bower L., Chung M.K., Goldsmith A.D., Jones T., Karlin-Neumann G., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M., Nguyen M., Palm C.J., Pham P.K., Quach H.L., Sakano H., Southwick A., Toriumi M., Yamada K., Yu G., Davis R.W., Theologis A., Ecker J.R.;

"Arabidopsis cDNA clones.";

Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K., Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bower L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kamaya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J., Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shim P., Southwick A., Shinohara K., Davis R.W., Ecker J.R., Theologis A.;

"Arabidopsis full length cDNA clones.";

Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.

-1-SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

EMBL: AF234990; AAG0342.1; -

EMBL: AY080753; AAL85999.1; -

HSP: P05713; 3RAB.

InterPro: IPR003579; GTPase\_Rab.

InterPro: IPR001806; Ras\_transfmg.

InterPro: IPR002078; S1954\_interact.

InterPro: IPR005225; Small\_GTP.

Pfam: PF00071; ras; 1.

PRINTS: PRO0449; RASTRNSFRMG.

SMART: SM00175; RAB; 1.

TIGRERMS: TIGR00231; small\_GTP; 1.

PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.

GTP-binding; Lipoprotein.

SEQUENCE 202 AA; 22313 MW; D0129A296C3EPDFC CRC64;

Query Match

Best Local Similarity 71.4%; Score 633; DB 10; Length 202;

Best Local Similarity 65.8%; Pred. No. 2.2e-43;

Matches 133; Conservative 10; Mismatches 25; Indels 34; Gaps 3;

DB 4 MNEPDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRIETLDEGKTILQ 63

DB 1 MNEPDYLFKLLIGDSGVKSCLLRFADDTYTESYSTIGVDFKRIETLDEGKTILQ 60

DB 64 I-----ESFNVMKQMLEIDRYASENVNKLIVG 91

DB 61 IMDTAGGERFRTTSSYRGAGHIIYVDTLSEFNVMKQMLEIDRYASENVNKLIVG 120

DB 92 NKCDLTKKVVDTTAKAFADSLGIPLETSAKNAITNVEOSFPMMAEIKRMGCGATAG 151

DB 121 NKCDLTSOKVSTETAKAFADSLGIPLETSAKNAITNVEAFMATAITRMA -SOPAG 179

DB 152 GAESNVKIQSTPVKQSGGCC 173

DB 180 GSKPPTVQIRGQPVNOQSG -GCCC 200

RESULT 15

024112 PRELIMINARY; PRT; 203 AA.

AC 024112; 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Small GTP-binding protein.

OS Nicotiana plumbaginifolia (Leadwort-leaved tobacco).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.

OC NCBI\_TaxID=4092;

OX [1]

RP SEQUENCE FROM N.A.

RA Borisjuk N., Stalio L., Kaydamov C., Senger S., Tewes A.,

RA Mantoufel R.;

RT "Cloning and characterization of cDNA clones differentially expressed

RT during somatic embryogenesis of Nicotiana plumbaginifolia: a mRNA

RT differential display approach."

RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

RL -1- SIMILARITY: TO RAS PROTEINS. BELONGS TO THE RAB SUBFAMILY.

CC EMBL: Y08425; CAA69701.1; -.

DR HSP; P01112; 1PL.

DR InterPro: IPR003579; GTPase\_Rab.

DR InterPro: IPR001230; Prenyl\_site.

DR InterPro: IPR001806; Ras\_trnsfmg.

DR InterPro: IPR002078; Slg54\_interact.

DR InterPro: IPR005225; Small\_GTP.

DR Pfam: PF00071; ras; 1.

DR PRINTS: PR00449; RASTRNSFRMG.

DR SMART: SM00175; Rab; 1.

DR TIGRFAMs: TIGR00231; small\_gtp; 1.

DR PROSITE: PS00294; PRENYLATION; UNKNOWN\_1.

DR PROSITE: PS00675; SIGMA54\_INTERACT\_1; UNKNOWN\_1.

KW GTP-binding; Lipoprotein.

KW GTP-binding; Lipoprotein.

SEQUENCE 203 AA; 22516 MW; 13799234FF719DE7 CRC64;

Query Match 71.4%; Score 632.5; DB 10; Length 203;  
Best Local Similarity 63.9%; Pred. No. 2.5e-43;  
Matches 129; Conservative 12; Mismatches 28; Indels 33; Gaps 2;

QY 4 MNPEDYLFKLLIGDSGVKSCLLRFADPTYESYSTIGVDPKIRTIELDGKTIKQ 63  
DB 1 MNPEDYLFKLLIGDSGVKSCLLRFADPTYESYSTIGVDPKIRTFVQDGTIKIQ 60

QY 64 I-----ESFNVKQWLQEIIDRYASENVKILYG 91  
DB 61 IWTAGGERFRTTSSYRGAGIIVYDVTQESFNNVKQWLSEIDRYASDSVKNILYG 120

QY 92 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEQSFMTAAEIKRKGPGATAG 151  
DB 121 NKCDLTKKVVDTTAKFADSLGIPLETSAKNATNVEQSFMTAAEIKRKGPGATAG 180

QY 152 GAESKSNVKGISTPVKQSGGGCC 173  
DB 181 NARPPVQIRGQPVNOK-SGCC 201

Search completed: November 17, 2002, 13:06:53  
Job time : 83 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame-plus\_p2n model

Run on: November 17, 2002, 13:06:58; Search time 1959 Seconds  
(without alignments)  
1430.229 Million cell updates/sec

Title: US-09-820-003a-2  
Perfect score: 886  
Sequence: 1 MSHNNPEXDYLEFKLLIGDS.....EKSNNKIGSPFKVQSGGCC 173

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+P2n.model -DEV-xih  
-O/cgn2.1/USPTO.spool/US09820003/runat\_13112002.134616.3189/app\_query.fasta.1.327  
-DB-EST -OPMT-fastap -SUFFIX-est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE-pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE-LOCAL  
-OUTFMT=ptc -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09820003 @cgn.1.1.899.8runat\_13112002.134616.3189 -NCPU=6 -ICPU=3  
-NO\_XLPY -NO\_MMAP -LARGOQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120  
-MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:\*  
1: em.estba:\*  
2: em.esthum:\*  
3: em.estlin:\*  
4: em.estm:\*  
5: em.estm:\*  
6: em.estm:\*  
7: em.estm:\*  
8: em.hic:\*  
9: gb.est1:\*  
10: gb.est2:\*  
11: gb.est3:\*  
12: gb.est4:\*  
13: gb.est5:\*  
14: gb.est6:\*  
15: em.estlum:\*  
16: em.estlum:\*  
17: gb.gss:\*  
18: em.gss.hum:\*  
19: em.gss.hum:\*  
20: em.gss.hum:\*  
21: em.gss.hum:\*  
22: em.gss.hum:\*  
23: em.gss.hum:\*  
24: em.gss.hum:\*  
25: em.gss.hum:\*  
26: em.gss.hum:\*  
27: em.gss.hum:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	860	97.1	948	9	AL539022
2	860	97.1	948	9	AL539022
3	849	95.8	1130	13	BM542833
4	847	95.6	862	14	BM223273
5	847	95.6	862	14	BM223273
6	831.5	93.8	646	13	BQ262634
7	825	93.1	646	13	BQ262634
8	822	92.8	646	13	BQ262634
9	820	92.6	646	13	BQ262634
10	819	92.4	646	13	BQ262634
11	814	91.9	646	13	BQ262634
12	809	91.3	646	13	BQ262634
13	808.5	91.3	642	13	BQ262634
14	807	91.1	642	13	BQ262634
15	806	91.0	642	13	BQ262634
16	801	90.4	642	13	BQ262634
17	800	90.3	642	13	BQ262634
18	800	90.3	642	13	BQ262634
19	795	89.7	642	13	BQ262634
20	795	89.7	642	13	BQ262634
21	794	89.6	642	13	BQ262634
22	794	89.6	642	13	BQ262634
23	788.5	89.0	642	13	BQ262634
24	788.5	89.0	642	13	BQ262634
25	788	88.9	642	13	BQ262634
26	787	88.8	642	13	BQ262634
27	783	88.4	642	13	BQ262634
28	783	88.4	642	13	BQ262634
29	780	88.0	642	13	BQ262634
30	778.5	87.9	642	13	BQ262634
31	777	87.7	642	13	BQ262634
32	777	87.7	642	13	BQ262634
33	776	87.6	642	13	BQ262634
34	775.5	87.5	642	13	BQ262634
35	775	87.5	642	13	BQ262634
36	769.5	86.9	642	13	BQ262634
37	769	86.8	642	13	BQ262634
38	768.5	86.7	642	13	BQ262634
39	768.5	86.7	642	13	BQ262634
40	768.5	86.7	642	13	BQ262634
41	768.5	86.7	642	13	BQ262634
42	768.5	86.7	642	13	BQ262634
43	768.5	86.7	642	13	BQ262634
44	768.5	86.7	642	13	BQ262634
45	768.5	86.7	642	13	BQ262634

# ALIGNMENTS

RESULT 1	AL539022	948 bp	mrna	linear	EST 16-FEB-2001
LOCUS	AL539022	948 bp	mrna	linear	EST 16-FEB-2001
DEFINITION	AL539022 LTR-FL013-FBcrn1 Homo sapiens cDNA clone CSDF030YG06 5				
ACCESSION	AL539022				
VERSION	AL539022.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				

JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 source  
 1. 948  
 Location/Qualifiers

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="F1013\_FBRn1"  
 /dev\_stage="pooled tissue from post conception fetuses (20 week, 24 week and 26 week)"  
 /lab\_host="DH10B"  
 /note="Organ: Fetal brain; Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the PCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com url : http://fulllength.invitrogen.com"

BASE COUNT 277 a 187 c 239 g 244 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,87e-94 Length: 948  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 Gaps: 1

US-09-820-003a-2 (1-173) x AL539022 (1-948)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 Db 166 ATGTCAGCATGATCCGCAATATGATTATTATTCAGTTACTTGTGATGGCGACTCA 225  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 Db 226 GGGGTGGGAAGTCTGCTCTTCTTAGGTTGCAATATACATATACAGAAAGCTAC 285  
 QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
 Db 286 ATCAGCACATTTGGTGTGATTTCAAAATAGAACTATAGATTAGACGGGAAACATC 345  
 QY 61 LysLeuGlnIle----- 64  
 Db 346 AACCTCAATATGAGACAGACAGCCAGAAAGATTTCGAACATCACCTCCAGTTAT 405  
 QY 65 -----GluSerPheAsn 68  
 Db 406 TACAGAGACCCCATGCGATCATAGTTGTATGATGACAGATCAGAGTCCCTTCAT 465  
 QY 69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyrIleAspSerGluAsnValAsnLysLeu 88  
 Db 466 AATGTTAAACAGTGGCTGCGCAAAATAGATCGTTATGCCGTAAGAAATGTCACAAATTTG 525  
 QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
 Db 526 TTGGTAGGAACAATGTGATCTGACCAACAAGAAAGTACTAGACTACACAGCGAAG 585  
 QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
 Db 586 GAATTTGCTATTCCCTTGGAAATTCCTTTTGGAAACAGTGTAAAGATGCAACCAAT 645  
 QY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyValA 148  
 Db 646 GTAGAACAGCTTTTCATGACGATGGCAGCTGAGATTAAAAAGCAATGGGTCCCGGAGCA 705

QY 149 ThrAlaGlyValAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
 Db 706 ACAGCTGGTGGTCTGTGAGAGTCCAAATGTTAAATTCAGAGCACTCCAGTCAAGCAGTCA 765  
 QY 169 GlyGlyGlyCysCys 173  
 Db 766 GGTGAGAGTTGCTGC 780

RESULT 2  
 BM919860 976 bp mRNA linear EST 12-MAR-2002  
 LOCUS BM919860  
 DEFINITION AGENCOURT\_6708189 NIH\_MGC\_120 Homo sapiens cDNA clone IMAGE:5749635  
 5', mRNA sequence.  
 ACCESSION BM919860  
 VERSION BM919860.1 GI:19370239  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
 1 (bases 1 to 976)  
 NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation (LLNL)  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLAM12779 row: f column: 04  
 High quality sequence stop: 757.

FEATURES  
 source  
 1. 976  
 Location/Qualifiers

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="5749635"  
 /clone\_lib="NIH\_MGC\_120"  
 /lab\_host="DH10B"  
 /note="Organ: Pooled pancreas and spleen; Vector: PCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (Rcovr site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH-MGC Library."

BASE COUNT 285 a 196 c 246 g 248 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.95e-94 Length: 976  
 Score: 860.00 Matches: 173  
 Percent Similarity: 84.39% Conservative: 0  
 Best Local Similarity: 84.39% Mismatches: 0  
 Query Match: 97.07% Indels: 32  
 Gaps: 1

US-09-820-003a-2 (1-173) x BM919860 (1-976)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
 Db 177 ATGTCAGCATGATCCGCAATATGATTATTATTCAGTTACTTGTGATGGCGACTCA 226  
 QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
 Db 237 GGGGTGGGAAGTCTGCTCTTCTTAGGTTGCAATATACATATACAGAAAGCTAC 296  
 QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60

Db		ATCACCAACAATTGTCGTGCATTTCCAAATAAGACTATGAGTACAGCGGAACAATC	64
Oy	61	LysLeugInIle-----	64
Db	357	AAGCTTCMAAATNtGGGACACAGCGGCCGAGAAGATTTTCGAACATCACCCTCACTT	416
Oy	65	-----GlusPheasn	68
Db	417	TACAGAGACCCCAGCATCATAGTTGGTGTAATGATGTACAGATCAGAGAGTCTTCAT	476
Oy	69	AsnValLysGIIntRpleungInGuIlleAspaRTyrAlaserGIuasbnValasnlYseu	88
Db	477	AATGTAAACAGGGGCTCGACAGAAATAGATCGTATGCCAGTGAATGCAACAAATTG	536
Oy	89	LeuValGLyasnlYscYsaSPleuthrThrlyslYsaValaIaSPryrrthrThralYls	108
Db	537	TTGGTAGGAAACAAATGATCATGACACCAAAAAGATGTAGACTACACACAGCAG	596
Oy	109	GIUpheaIaasPserLeugLyIlePopeHeuclunThrSerAlalYSaSanaIathrasn	128
Db	597	GAAATTCCTCATATCCCTGGAAATCCGTTTTTGGAAACAGTCAGTAAGATGCACGAT	656
Oy	129	ValGIuInsErPhemeThrMetAlaIaLuIleLyLaLYsaQmteGIYPooGIyla	148
Db	657	GTAGAACAGCTTCATGACAGGACGACGTGATTAACCGAATGGGTCCCGSAGCA	716
Oy	149	ThrAlaGIyAlaGIuLYssERanValylsIleGIlnSerThrProVALysGInsEr	168
Db	717	ACAGCTGCTGTCGTGAGAAAGTCCATGTAAATTCAAGACACTCAAGCAAGCAGCA	776
Oy	169	GIyGIyGIyXcYScYs	173
Db	777	GCTGAGAGTTGCTGC	791
RESULT 3			
LOCUS	BMS42833	1130 bp	mRNA linear EST 20-FEB-2002
DEFINITION	AGENCOUPT_6426035 NIH_MGC_71	Homo sapiens cDNA clone IMAGE:5520670	
ACCESSION	BMS42833		
VERSION	BMS42833.1	GI:18772697	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
JOURNAL	1 (bases 1 to 1130)		
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov/ Plate: LHAM12185 row: a column: 23 High quality sequence stop: 597. Location/Qualifiers 1..1130 organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="5520670" /clone_lib="NIH_MGC_71" /tissue_type="Telomerasecoma" /lab_host="DH10B (phage-resistant)" /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT, Average insert size 2.1 kb."		

BASE COUNT	319 a	241 c	290 g	280 t
ORIGIN				
Alignment Scores:				
Prod. No.:	5.4e-93	Length:	1130	
Score:	849.00	Matches:	171	
Percent Similarity:	83.41%	Conservative:	0	
Best Local Similarity:	83.41%	Mismatches:	2	
Query Match:	95.82%	Indels:	32	
	13	Gaps:	1	
US-09-820-003A-2 (1-173) x BM542833 (1-1130)				
Oy 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20				
Db 178 ATGTCGACGATGAAATCCCGAATATGATATTTATATCAAGTCCTTCGATGGCCACTCA 237				
Oy 21 G1yAlG1yLysSerCysLeuLeuLeuAArgPheAlaAspSerThrTyrTrgIuseTyr 40				
Db 238 GGGGTGGAAAGCTTGCGCTTCCTTCTTACGATTCACATGATACATATACGAAAGCTAC 297				
Oy 41 IleserThrIleG1yAlAspPheLysIleAthrIleG1uLeuAAspGlyLysThrIle 60				
Db 298 ATCACAACAATTCGTGTGATTTCAATAATAGAACTATGAGAGTGAAGCGGAAACATC 357				
Oy 61 LysLeuGlnIle----- 64				
Db 358 AAGCTCAATATAGGGACACAGACGCGCAAGAAATTCGAAACATCACCCTCCAGTTAT 417				
Oy 65 -----G1usePheAsn 68				
Db 418 TACAGAGAGACCCATGCATCATATGTTGTTATGATGTGACATGCACATCGAGTCTTCAT 477				
Oy 69 AsnValLysGlnTrpLeuGlnIuIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88				
Db 478 AATGTTAAACAGGCGCTCGACGAAATAGATCGTTATGCCAGTGAATATGCAACAAATGG 537				
Oy 89 LeuValG1yAsnLysCysAspLeuThrIleLysLysValValAspTyrThrIleLys 108				
Db 538 TTGGTAGGGAACAAATGTGATCTGACCAACAAAGAAATAGTAGTACACACACGCGAAG 597				
Oy 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluIleSerAlaLysAsnAlaThrAsn 128				
Db 598 GAATTTCTCATATCCCTTGGAAATTCGTTTGGAAACGACGCTGTAAGAAATGCACGAT 657				
Oy 129 ValGluGlnSerPheMetThrMetAlaIleGluIleLysLysArgMetGlyProGlyAla 148				
Db 658 GTAGAAACGCTTCATGACAGATGAGAGCTGATTTAAACGAAATGGGTCCCGGACCA 717				
Oy 149 ThrAlaGlyAlaG1uLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168				
Db 718 ACAGCTGTGTGTCTGAGAAATGCATGTTAAATTCAGAGCACTCCATGCAAGCATTC 777				
Oy 169 GlyGlyG1yCysCys 173				
Db 778 ACTGGAGTGTCTGC 792				
RESULT 4				
LOCUS BQ223273	862 bp	mRNA	linear	EST-02-MAY-2002
DEFINITION AGNCNOUT_7518264 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6066561				
ACCESSION BQ223273				
VERSION BQ223273.1				
KEYWORDS GI:20404673				
SOURCE EST.				
ORGANISM human.				
REFERENCE Eukarya: Metazoa: Chordata: Cranata: Vertebrata: Euteleostomi:				
AUTHORS Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.				
TITLE NIH-MGC http://mgi.nci.nih.gov/				
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)				
UNPUBLISHED (1999)				

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM13344 row: k column: 10  
High quality sequence stop: 728.  
Location/Qualifiers

## FEATURES

source

1.862  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:606561"  
/clone\_1ib="NIH\_MGC\_92"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Testis; Vector: pCMV-SPORT6; Site.1: NotI;  
Site.2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC Library."

BASE COUNT 254 a 170 c 228 g 210 t  
ORIGIN

## Alignment Scores:

Pred.: No.: 6.26e-93 length: 862  
Score: 847.00 Matches: 173  
Percent Similarity: 83.98% Conservative: 0  
Best Local Similarity: 83.98% Mismatches: 0  
Query Match: 95.60% Indels: 33  
Gaps: 14

US-09-820-003a-2 (1-173) x BQ23273 (1-862)

QY 1 MetSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
DB 179 ATGTCAGCATGATCCGGAATATGATTATTTCAAGTACTTGCATGGCGACTCA 238  
QY 21 GlyAlaGlySerCysLeuLeuLeuAlaPheAlaAspThrTyrThrGluSerTyr 40  
DB 239 GGGGTGGGAAGCTTCCCTCTTCTTGGTGCAGATATACATATACAGAAAGCTAC 298  
QY 41 ILSerThrIleGlyValaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 299 ATCAGCACAATTGGTGTGATTTCAAAATAGAACTATAGAGTAGACGGGAAACATC 358  
QY 61 LysLeuGlnIle----- 64  
DB 359 AAGCTTAAATATGGACACAGCAGCGCCAGGAAAGATTGGAACAATCACCCTCAATT 418  
QY 65 -----GluSerPheAsn 68  
DB 419 TACAGAGACCCCATGCGATCATAGTTGTATGATGTCAGATCAGATGCTTCAAT 478  
QY 69 AspValysGlnTrpLeuGlnIleAspArgTyrAlaSerGluAsnValaAsnLysLeu 88  
DB 479 AATGTTAAACAGTGGCGCAGGAATATGATGCTTATCCAGTGAATGTCACAAATTTG 538  
QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValaValaAspTyrThrAlaLys 108  
DB 539 TTGGTAGGAACAATGTGATCTGACCAACAAGAACTAGTACGCTACACACAGCGAAG 598  
QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
DB 599 GAATTTGCTATTCCTCTGGAATTCCTTTTGGAAACAGTGTAGAAATGCAACGAAT 658  
QY 129 ValGluGlnSerPheMetThrMetAlaAlaGlnIleLysLysArgMetGlyProGlyAla 148  
DB 659 GTAGAACAGCTTTCATGAGATGCGAGCTGAGATTAAAGGCAATGGGTCCGAGACA 718

QY 149 ThrAlaGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
DB 719 ACAGCTGGTGGCTCGTAGAGATCCAAATGTTAAATTCAGACACTCCAGTCAAGCAGCTCA 778  
QY 169 Gly-GlyGlyCysCys 173  
DB 779 GGTGGAGGTGCTGC 794

## RESULT 5

BQ926234

LOCUS

DEFINITION

AGENCOURT\_8801639 NIH\_MGC\_129 Mus musculus CDNA clone IMAGE:6315467

5', mRNA sequence.

BQ926234.1 GI:22341265

VERSION

KEYWORDS

SOURCE

ORGANISM

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 944)

NIH-MGC <http://mgc.nci.nih.gov/>

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: Susan L. Sullivan, PhD.

CDNA Library Preparation: Resgen, Invitrogen Corp

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13745 row: n column: 12

High quality sequence stop: 750.

Location/Qualifiers

1.944

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="IMAGE:6315467"

/clone\_1ib="NIH\_MGC\_129"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: Olfactory epithelium; Vector:

pCMV-SPORT6.1.cdb; Site.1: EcoRV; Site.2: NotI; Cloned

unidirectionally. Primer: Oligo dT. Average insert size

2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: this

is a NIH-MGC Library."

BASE COUNT 282 a 182 c 211 g 266 t 3 others

ORIGIN

## Alignment Scores:

Pred. No.: 7.2e-93 length: 944  
Score: 847.00 Matches: 170  
Percent Similarity: 84.16% Conservative: 0  
Best Local Similarity: 84.16% Mismatches: 0  
Query Match: 95.60% Indels: 32  
Gaps: 14

US-09-820-003a-2 (1-173) x BQ926234 (1-944)

QY 4 MetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSerGlyValGly 23  
DB 20 ATGAATCCGGAATATGATTATTTATCAAGTTACTTCTGANTGGCGATTTCTGGGGTTGA 79  
QY 24 LysSerCysLeuLeuLeuAlaGpPheAlaAspAspThrTyrThrGluSerTyrIleSerThr 43  
DB 80 AAGTCTGCTTCTCTCTAGGTTTGCAGATGATACGATACGGAACCTCATCAGCACA 139  
QY 44 ILleGlyValaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGln 63  
DB 140 ATTGGTGTGATTTCAAGATACGACTATAGAGTTAGATGGGAAACAAACAAAGCTACAG 199  
QY 64 Ile----- 64

```

Db 200 ATATGGACACAGACAGCCAGGAAAGATTTCAGACATCTCCAGTTATTACAGAGCA 259
QY 65 -----GUSERPHEASNAVALYS 71
Db 260 GCCCATGCATCATAGTGTGTATGATGACAGATCAGAGCTCTTCATTAATAAGCTTAAA 319
QY 72 GlnTrpLeuGlnIuIleAspArgTyrAlaSerGluAsnValAsnLysLeuLeuValGly 91
Db 320 CAGTGGCTCAGAGATAGATCGCTACGCCAGTGAATGCTCAACAAAGCTGTGGTGGG 379
QY 92 AsnLysCysAspLeuThrThrLysValValAspTyrThrThrAlaLysGluPheAla 111
Db 380 AACCAATGATACCTGACCAACAAAGAGTAGACTACAAACAGCAAGAGGATTTGCA 439
QY 112 AsperLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln 131
Db 440 GATTCCTTGATTCATCTTTTGGAAACAGTCTAAGAACCAAGATGTAGAACAG 499
QY 132 SerPheMetThrMetAlaAlaGluIleLysLysAlaGlyProGlyAlaThrAlaGly 151
Db 500 TCTTTCATGACGATGCGACGTGACATTAAGCAATGGCTCTGACCTCAGCTGTGT 559
QY 152 GlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGlyGly 171
Db 560 GGTCCGAGAGATCCATGTTAAATCCAGACACTCCAGTCAAGCAGTCAAGTGGAGGC 619
QY 172 CysCys 173
Db 620 TGCTGC 625

RESULT 6
LOCUS B0219698 646 bp mRNA linear EST 02-MAY-2002
DEFINITION B0219698 NIBB Mochii normalized Xenopus early gastrula library
ACCESSION B0219698
VERSION B0219698.1 GI:17592487
KEYWORDS EST
SOURCE B0219698.1
ORGANISM B0219698.1
Xenopus laevis
Pukateya, Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 646)
AUTHORS Kitayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-I, T. and Kohara,
T.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasi Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
source
1..646
Location/Qualifiers
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_id="NIBB Mochii normalized Xenopus early gastrula
library"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/note="Vector: pBSRN3; Site_1: NotI; Site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Newkooop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
BASE COUNT 208 a 129 c 161 g 148 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 3,16e-91 Length: 646
Score: 831.50 Matches: 169
Percent Similarity: 82.93% Conservative: 1
Best Local Similarity: 82.44% Mismatches: 2
Query Match: 93.85% Indels: 33
DB: 13 Gaps: 2

US-09-820-003a-2 (1-173) x B0219698 (1-646)
QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20
Db 35 ATGTCACACATGATCCGAGATGATGATTAATTAATTAATTAATTAATTAATTAATTA 94
QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspTyrThrTrpGluSerTyr 40
Db 95 GGTGGGGAAGCTTGTCTCTGCTAGAGTTTGAGATGACATCATCAGACAAAGATTAC 154
QY 41 IleSerThrIleGlyValAlaAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60
Db 155 ATCAGTACCAATGTGTGTTGATTTTAAATCAGACTATATAGATTAGATGAAAAACATTC 214
QY 61 LysLeuGlnIle----- 64
Db 215 AAACGTCAATTTGGGACACCTGCGCGCAGGAAAGATTTCGAACCATTAATCAAGTTAC 274
QY 65 -----GUSERPHEASN 68
Db 275 TACAGGGAGCCGACGCGCATCTGTTGTGTATGACGTACAGACGACGATCTTCAAC 334
QY 69 AsnValLysGlnTrpLeuGlnIuIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88
Db 335 AATGTCACAGATGCTGCTTCAAGAAATAGATCTGATGCAATGCAATGCAATGCAATGCA 394
QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysValValAspTyrThrThrAlaLys 108
Db 395 TTGGTAGCACAACAATGATGATCTCCACAAAGAAAGTTGTGGACTCCACACAGCAAAAG 454
QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128
Db 455 GAGTTTGCAGACTCTCTGGGAATTCATTTTGGAAACAGCGCAAGAAAGCCGCAAAAC 514
QY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysAlaGlyProGlyAla 148
Db 515 GTTAGAGCAGGCGCTTCATGACAAATGACAGCGAGATCAAAAAGCAATGGCGCTGCGCC 574
QY 149 ThrAlaGlyIleAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168
Db 575 ACAGAGGCGGTCAAGAGAAAG---AACGTCAAAATCCAGAGCACTCCAGTCAAGAGTCT 631
QY 169 GlyGlyGlyCysCys 173
Db 632 AGTGGAGGATGCTGC 646

RESULT 7
LOCUS B0219698 957 bp mRNA linear EST 02-MAY-2002
DEFINITION B0219698 IMAGE:6051445 5', mRNA sequence.
ACCESSION B0219698
VERSION B0219698.1 GI:20401098
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC http://mgi.mc.nhl.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.

```

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM13305 row: e column: 14  
High quality sequence stop: 671.

## FEATURES

source

1. .957

Location/Qualifiers

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6051445"  
/clone\_lib="NCI\_CGAP-St1"  
/lab\_host="DH10B (711-resistant)"  
/note="Organ: Stomach; Vector: pCMV-SPORT6; Site: 1: SalI;  
Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 1.77 kb. Library constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."  
BASE COUNT 269 a 208 c 250 g 229 t 1 others  
ORIGIN

## Alignment Scores:

Pred. No.:	3,59e-90	Length:	957
Score:	825.00	Matches:	171
Percent Similarity:	83.41%	Conservative:	0
Best Local Similarity:	83.41%	Mismatches:	2
Query Match:	93.12%	Indels:	33
DB:	14	Gaps:	1

US-09-820-003a-2 (1-173) x BQ219698 (1-957)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
|||||  
DB 200 ATGTCCAGCATGATCCCGAATATGATTATTATTCAGTTACTTCTGATTGGCGATTCT 259  
QY 21 GlyValGlySerSerCysLeuLeuLeuArgPheAlaAspSphTrpThrGluSerTyr 40  
|||||  
DB 260 GGGGTGGAAAGTCTCCCTCTCTTACGATGATACGATACGATACGATACGATACGATAC 319  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
|||||  
DB 320 ATCAGCACAATTGGTGTGATTTCACATACGACATACGATACGATACGATACGATACGATAC 379  
QY 61 LysLeuGlnIle----- 64  
|||||  
DB 380 AAGCTACAGATATGGACACAGCAGCCAGAAAGATTGCAACATCCTCCAGTTAT 439  
QY 65 -----GluSerPheAsn 68  
|||||  
DB 440 TACAGAGAGCCCATGGCATCATAGTTGTATGATGATGACAGATCAGAGTCCCTCAAT 499  
QY 69 AsnValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
|||||  
DB 500 AAGCTTAAACAGTGGCTGCGACGAGATGATGCTACGCCGATGAATATGTCACAACTTG 559  
QY 89 LeuValGlyAsnLysCysAspLeuThrThrLysLysValValAspTyrThrThrAlaLys 108  
|||||  
DB 560 TTGGTAGGGAACAATGTGACCTGACCAACAAGAAAGTACTAGACTACACAGCAAG 619  
QY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
|||||  
DB 620 GAATTTCACATTCCTCTGGAAATTCATTTTGGAAACCGTCTAAGAACGCAACGAT 679  
QY 129 ValGluInserPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
|||||  
DB 680 GTTGAACAGCTTTTCATGACGATGCGAGCTGAGATTANAAAGCAATGGGTCTCGGAGCT 739  
QY 149 ThrAlaGlyLysAlaGluLysSerAsnValLysIleGluInserThrProValLysGlnSer 168  
|||||  
DB 740 ACAGCTGGTGGTGGCGAAGAAATCAT- GTTAAATCCAGAGCACTCCAGTCAAGCAATCA 798  
QY 169 GlyGlyLysCysCys 173

|||||  
DB 799 GGTGAGGCTGCTGC 813

## RESULT 8

BQ437146

LOCUS

## DEFINITION

BQ437146 864 bp mRNA linear EST 24-MAY-2002  
AGENCOURT\_7909781 NIH\_MGC\_71 Homo sapiens cdna clone IMAGE:6156281  
5', mRNA sequence.

## ACCESSION

BQ437146

BQ437146.1 GI:21176210

## VERSION

EST.

## KEYWORDS

human.

## ORGANISM

Homo sapiens

## REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 864)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapps-r@mail.nih.gov](mailto:cgapps-r@mail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM13500 row: e column: 18

High quality sequence stop: 658.

## FEATURES

source

Location/Qualifiers

1. .864

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6156281"  
/clone\_lib="NIH\_MGC\_71"  
/tissue\_type="Jellyfish sarcoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: NotI;  
Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
Average insert size 2.1 kb."  
BASE COUNT 249 a 171 c 231 g 213 t  
ORIGIN

## Alignment Scores:

Pred. No.:	7.15e-90	Length:	864
Score:	822.00	Matches:	170
Percent Similarity:	82.93%	Conservative:	0
Best Local Similarity:	82.93%	Mismatches:	2
Query Match:	92.78%	Indels:	33
DB:	14	Gaps:	1

US-09-820-003a-2 (1-173) x BQ437146 (1-864)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
|||||  
DB 184 ATGTCCAGCATGATCCCGAATATGATTATTATTCAGTTACTTCTGATTGGCGACTCA 243  
QY 21 GlyValGlySerSerCysLeuLeuLeuArgPheAlaAspSphTrpThrGluSerTyr 40  
|||||  
DB 244 GGGGTGGAAAGTCTCCCTCTCTTACGATGATACGATACGATACGATACGATACGATAC 303  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
|||||  
DB 304 ATCAGCACAATTGGTGTGATTTCACAAATTAAGAACTTACAGTTAGCGGAAACATC 363  
QY 61 LysLeuGlnIle----- 64  
|||||  
DB 364 AAGCTTCAATATATGGACACAGCAGCCAGAAAGATTGCAACATCACCCTCAGTTAT 423  
QY 65 -----GluSerPheAsn 68  
|||||  
DB 424 TACAGAGAGCCCATGGCATCATAGTTGTATGATGATGACAGATCAGAGTCCCTCAAT 483

Mon Nov 18 08:42:06 2002

us-09-820-003a-2.fst

Page 7

OY 69 AasnVallysgIntRpleuGlnGluLeasparyTyrAlasergLysanValasnLysIeu 88  
DB 484 AATGTTAAACAGTGGCTGCAGGAAATAGATCTTATGACGAGAAATGTCACAAATG 543  
OY 89 LeuValGlyAasnLysCysAspLeuThrThrLysValValAspTyrThrThrAlaLys 108  
DB 544 TTGGTAGGAGCAAAATGATGATCTGCACCAAGAAAGTAGTAGTACACACAGAG 603  
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
DB 604 GAATTTGCTGATCCCTTGGAATCCGTTTGGAAACAGAGCTAGACATGCAAGAT 663  
OY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
DB 664 GTAGAACAGCTCTTCATGACGATGAGCTGATTAAGAAAGCAATGGTCCGAGCA 723  
OY 149 ThrAlaGlyAlaGlyAlaGlyLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
DB 724 ACAGCTGGTGGCTGCAGGAAAGTCAATGGTTAAATTCAGAGCTCCAGTCAAGCACTCA 783  
OY 169 Gly-GlyGlyCys 172  
DB 784 GGTGGAGGTTGC 796  
RESULT 9 901 bp mRNA linear EST 21-AUG-2002  
LOCUS B0944368  
DEFINITION IMAGE:6441874 5', mRNA sequence.  
ACCESSION B0944368  
VERSION B0944368.1 GI:22359846  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 901)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM13965 row: m column: 11  
High quality sequence stop: 662.  
Location/Qualifiers  
1. 901  
FEATURES  
source  
/organism="Mus musculus"  
/strain="FVB/N-3"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6441874"  
/clone\_lib="NCI-CGAP Mam2"  
/tissue\_type="tumor, biopsy sample"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; site\_1: SalI;  
site\_2: NotI; Cloned by Life Technologies. Investigator  
Library constructed: Gilbert Smith, NIH  
providing samples: Gilbert Smith, NIH  
BASE COUNT 246 a 197 c 243 g 214 t  
ORIGIN  
Alignment Scores: 1,34e-89 length: 901  
Pred. No.: 820.00 Matches: 170  
Score: 82.61% Conservative: 1  
Percent Similarity:

Best Local Similarity: 82.13% Mismatches: 2  
Query Match: 92.55% Indels: 34  
DB: 14 Gaps: 1  
US-09-820-003a-2 (1-173) x B0944368 (1-901)  
OY 1 MetSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
DB 188 ATGTCAGATCAATCCGAAATGATTTATTATTCAGTACTCTGATGGCGATTCT 247  
OY 21 GlyValGlyLysSerCysLeuLeuAsnPheAlaAspAspThrTyrThrGluSerTyr 40  
DB 248 GGGGTGGAAAGTCCGCTTCCTTCTAGCTTCAGATCATGATACGAAACCTAC 307  
OY 41 TleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAsnGlyLysThrIle 60  
DB 308 ATCAGCAAAATGGGTGGATTCATTCAGATTCAGAACTATAGATTAGGGAACAAATC 367  
OY 61 LysLeuGlnIle-----GluSerPheAsn 64  
DB 368 AACCTACAGATATGGGACACACAGCGCAGGAAAGATTTCGAACATCACTTCCAGTTAT 427  
OY 65 -----GluSerPheAsn 68  
DB 428 TACAGAGAGCCATGCGATCATAGTTGTATGATGTGACAGATCAGAGATGCTTCAAT 487  
OY 69 AasnVallysgIntRpleuGlnGluLeasparyTyrAlasergLysanValasnLysIeu 88  
DB 488 AATGTTAAACAGTGGCTGCAGGAAATAGATCTGACGAGAAATGTCACAAAGTTG 547  
OY 89 LeuValGlyAasnLysCysAspLeuThrThrLysValValAspTyrThrThrAlaLys 108  
DB 548 TTGGTAGGAGCAAAATGATGATCTGCACCAAGAAAGTAGTAGTACACACAGCAAG 607  
OY 109 GluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsn 128  
DB 608 GAATTTGCTGATCCCTTGGAATCCGTTTGGAAACAGAGCTAGAACGCAAGAT 667  
OY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
DB 668 GTAGAACAGCTCTTCATGACGATGAGCTGATTAAGAAAGCAATGGTCCGAGCA 723  
OY 149 ThrAlaGlyAlaGlyAlaGlyLysSerAsnValLysIleGlnSerThrProValLysGln 168  
DB 728 ACAGCTGGTGGCTGCAGGAAAGTCAATGGTTAAATTCAGAGCTCCAGTCAAGCACTCA 787  
OY 168 ergLylGlyGlyCys 173  
DB 788 CAGGTGGAGGCTGCTGC 804  
RESULT 10  
LOCUS BM452262 1105 bp mRNA linear EST 05-FEB-2002  
DEFINITION IMAGE:6386191 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:552635  
ACCESSION BM452262  
VERSION BM452262.1 GI:18501302  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1105)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC/DCID/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/MLN at:  
http://image.lnl.gov

Plate: LHAM2200 row: j column: 12  
High quality sequence stop: 733.

## FEATURES

Location/Qualifiers  
1..1105

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1db="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-Sport6, site\_1: NotI;  
site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 320 a 216 c 279 g 282 t 8 others  
ORIGIN

## Alignment Scores:

Pred. No.: 2.43e-89 Length: 1105  
Score: 819.00 Matches: 172  
Percent Similarity: 83.50% Conservative: 0  
Best Local Similarity: 83.50% Mismatches: 1  
Query Match: 92.44% Indels: 34  
DB: 13 Gaps: 1

US-09-820-003a-2 (1-173) x BM452262 (1-1105)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
DB 182 ATGTCAGACGATGATCCCGAATGATGATTTATTCAGTACTTCGATGGCGACTCA 241  
QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
DB 242 GGGGTGGAAAGCTTGGCTTCTTCTTAGGTTTCAGATGATACATATACGAAAGCTAC 301  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 302 ATCAGCACAATGTTGGTGGATTCGAAATAGAACTATAGATGAGACGGGAAAAATC 361  
QY 61 LysLeuGlnIle----- 64  
DB 362 AAGCTTCAAAATATGGACACAGACAGGCGGAAAGATTTCGACAAATCACCCTCAGTTAT 421  
QY 65 -----GluSerPheAsn 68  
DB 422 TACAGAGAGCCCATGGCATCATAGTTGTGATGATGACAGATCAGGAGTCCCTTCAT 481  
QY 69 AsnValLysGlnTyrPheGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
DB 482 AATGTTAAACAGTGGTGCAGGAATATGATGCTATGCAATGAAATGCAACAATTTG 541  
QY 89 LeuValGlyAsnLysCysAspLeuThrLysLysValValAspTyrThrAlaLys 108  
DB 542 TTGGTAGGGAACAATGTGATCTGACCAACAAGAAAGTAGAGACTACACAACGCGAAG 601  
QY 109 GluPheAlaAspSerLeuGlyIleProPheGlnGluThrSerAlaLysAsnAlaThrAsn 128  
DB 602 GAATTTGCTGATTCCTTGGAAATTCGTTTGGAAACCAAGTGTAAAGATGCAACGAT 661  
QY 129 ValGluGlnSerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAla 148  
DB 662 GTAGAACAGTCTTCATGAGAGATGACAGCTGAGATTAAAGGAAAGGAGGCGGAGCA 721  
QY 149 ThrAlaGlyGlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSer 168  
DB 722 ACAGCTGTGTGCTGAGAA-TCCAAATGTTAAATTCAGAGCACTCCAGTCAACAGATCA 780  
QY 169 -GlyGlyGlyCysGys 173  
DB 781 NGGTGAGAGTTGCTGC 796

RESULT 11  
AM954378 690 bp mRNA linear EST 01-JUN-2000  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AM954378  
EST366448 MAGE resequences, MAGE Homo sapiens cDNA, mRNA sequence.  
AM954378  
AM954378.1 GI:8144061  
EST.  
human.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 690)  
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt  
/I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and  
Quackenbush,J.  
Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
Unpublished (2000)  
CONTACT  
Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@ligr.org  
Plate: 63  
Seq primer: Reverse.

## FEATURES

Location/Qualifiers  
1..690

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1db="MAGE resequences, MAGE"  
/note="Vector: BluescriptSKm"

BASE COUNT 219 a 130 c 168 g 173 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.81e-89 Length: 690  
Score: 814.00 Matches: 169  
Percent Similarity: 83.25% Conservative: 0  
Best Local Similarity: 83.25% Mismatches: 1  
Query Match: 91.87% Indels: 33  
DB: 10 Gaps: 1

US-09-820-003a-2 (1-173) x AM954378 (1-690)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPheLysLeuLeuIleGlyAspSer 20  
DB 31 ATGTCAGACGATGATCCCGAATGATGATTTATTCAGTACTTCGATGGCGACTCA 90  
QY 21 GlyValGlyLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
DB 91 GGGGTGGAAAGCTTGGCTTCTTCTTAGGTTTCAGATGATACATATACGAAAGCTAC 150  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
DB 151 ATCAGCACAATGTTGGTGGATTCGAAATAGAACTATAGATGAGACGGGAAAAATC 210  
QY 61 LysLeuGlnIle----- 64  
DB 211 AAGCTTCAAAATATGGACACAGCAGCGGAAAGATTTCGACAAATCACCCTCAGTTAT 270  
QY 65 -----GluSerPheAsn 68  
DB 271 TACAGAGAGCCCATGGCATCATAGTTGTGATGATGACAGATCAGGAGTCCCTTCAT 330  
QY 69 AsnValLysGlnTyrPheGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeu 88  
DB 331 AATGTTAAACAGTGGTGCAGGAATATGATGCTATGCAATGAAATGCAACAATTTG 390  
QY 89 LeuValGlyAsnLysCysAspLeuThrLysLysValValAspTyrThrAlaLys 108  
DB 391 TTGGTAGGGAACAATGTGATGATGACCAACAAGAAAGTAGAGACTACACAACAGGAGAG 450



Mon, Nov 18 08:42:06 2002

us-09-820-003a-2.1st

Page 9

```

OY 109 GIUPEALAAASPERLEUGLYILEPROPHLEUGLUTHRISERIALALYSASNAATHRASN 128
DB 451 GAAATTCCTGATCCCTTGGAATTCCTTTTGGAAACAGCTGCTAGAAATCAACGAAAT 510
OY 129 VALGIUGINSERPHETHEMERICALAGLUILELVSARWGETGLYPROGLYALA 148
DB 511 GTAGAACAGCTTTCATGACGATGCGACGATTAATAAAAGCAAGGATCCCGGACCA 570
OY 149 THRAGAGLYGALAGLUISERASVALYSILEGINSETRHR-PROVALYSGINSE 168
DB 571 ACAGCGGTGCTGCTGGAATAATCCAAATGTAATAATCCAGACACTCCAGTCAGAGTC 630
OY 168 RGIYGLY 170
DB 631 AGGTGGG 637

RESULT 12
AL530265 945 bp mRNA linear EST 13-FEB-2001
LOCUS AL530265 LTI_NFL001_NB04 Homo sapiens cDNA clone CS0DD009YH19 5
DEFINITION prime, mRNA sequence.
ACCESSION AL530265
VERSION AL530265.1 GI:12793758
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 945)
AUTHORS Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..945
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="LTI_NFL001_NB04"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Feng liang life
technologies, a division of 20850 USA Fax : (1) 301 610
8371 Email : fliang@life-techn.com"
http://fulllength.invitrogen.com"

BASE COUNT 280 a 186 c 224 g 246 t 9 others
ORIGIN

Alignment Scores:
Pred. No.: 3.19e-88 length: 945
Score: 809.00 matches: 171
Percent Similarity: 83.01% mismatches: 0
Best Local Similarity: 83.01% indels: 34
Query Match: 91.31% gaps: 1
DB: 9

US-09-820-003a-2 (1-173) x AL530265 (1-945)
OY 1 MetSerSerMetAsnProGluThrAspTyrLeuPheLysLeuLeuLeuLeuLysPse 20
DB 121 ATGTCCACGCAAGTAATCCAAATGATGATTATTATTCAGAGTACTTGATGCGACATCA 180
OY 21 G1YVALG1LYSERCYSLEULEUENAVRPHALASAPSPRTHRYTHRG1UBERTYR 40

DB 181 GGGCTGGGAAGAGTCTGCTTCCTTTAGGTTTCAGAGTACATCATATACAGAAAGCTAC 240
OY 41 ILESERTHRIIEGLYVALASPHLYSLIETATGTHRIIEGLULEUASPGLYSTHRIIE 60
DB 241 ATAGACACAATGGGTGGATTTCAAAATAGACATRTAGAGTTAGCGGGAAACAATTC 300
OY 61 LYSLEUGLIIIE----- 64
DB 301 AACCTCAAAATATGGGACACAGCAGGCCAGAAAGATTTCGAACATCCTCAGATTAT 360
OY 65 ----- GluSerPheAsn 68
DB 361 TACAGAGAGCCCATGCAATCATGTTGTGATGATGTGACAGATCAAGATCCTTCATAT 420
OY 69 ASNAVALLYSGINTPRLEUGLNGLUILEASPARGTYRILASERGLVASNAVALLYSLIEN 88
DB 421 AATGTTAAACAGTGTCTGACGAAATAGATCGTTTATGCCAGTGAATAATGTCACAAATTC 480
OY 89 LEUVALG1LYASALYSCYASPLEUTRTHRILYSVALYVALASPTYRTHR-THRALALYS 108
DB 481 TTGCTAGGGAACAATAGTATCTGACCAAGAAAGTAGAGTACACACAGCGCAG 540
OY 109 GIUPEALAAASPERLEUGLYILEPROPHLEUGLUTHRISERIALALYSASNAATHRASN 128
DB 541 GAAATTCCTGATCCCTTGGAATTCCTTTTGGAAACCACTGCTAAGAAATCAACGAAT 600
OY 129 VALGIUGINSERPHETHEMERICALAGLUILELVSARWGETGLYPROGLYALA 148
DB 601 GTAGAACAGCTTTCATGACGATGCGACGATTAATAAAAGCAAGGATCCCGA-GCA 659
OY 149 THRAGAGLYGALAGLUISERASVALYSILEGINSETRHR-PROVALYSGINSE 168
DB 660 ACAGCGGTGCTGCTGGAATAATCCAAATGTAATAATCCAGACACTCCAGTCAGAGTCA 719
OY 169 -G1YGLYGLYCYCS 173
DB 720 GGGTGGAGGTCTCTGC 735

RESULT 13
LOCUS B073109 642 bp mRNA linear EST 11-DEC-2001
DEFINITION B073109 NIBB Mochii normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL108P01 5', mRNA sequence.
ACCESSION B073109
VERSION B073109.1 GI:17503298
KEYWORDS EST.
ORGANISM African clawed frog.
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodidae; Xenopus.
1 (bases 1 to 642)
K14yAm, A., Teratsuka, C., Mochii, M., Ueno, N., Shin-I, T. and Kohara
Y.
REFERENCE 1
EXpressed genes in X. laevis embryo
AUTHORS
TITLE
JOURNAL
COMMENT
CONTACT: Tadasi Shin-I
National Institute of Genetics
Center for Genetic Resource Information
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..642
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_id="XL108P01"
/clone_lib="NIBB Mochii normalized Xenopus tailbud
library"
/tissue_type="whole embryo"
/dev_stage="stage 25"

FEATURES
source

```

/note="Vector: pBSRN3; Site\_1: NotI; Site\_2: EcoRI; CDNAS were Oligo-dT primed and directionally cloned. Staging according to Nieuwkoop and Faber. Library is substracted and constructed by N. Garrett and A.M. Zorn, (Wellcome/CRC Institute)."

BASE COUNT 212 a 137 c 150 g 142 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 2,03e-88 Length: 642

Score: 808.50 Matches: 164

Percent Similarity: 81.68% Conservative: 1

Best Local Similarity: 81.19% Mismatches: 4

Query Match: 91.25% Indels: 33

Gaps: 2

US-09-820-003a-2 (1-173) x BJ073109 (1-642)

QY 4 MetAsnProGluTyrAspTyrLeuPhelLysLeuLeuIleGlyAspSerGlyValGly 23  
|||||  
DB 10 ATGAAATCCCGAATATGACTATTATTCAAATTACTTTGATGGAGACTGTGTGGGA 69  
QY 24 LysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyrIleSerThr 43  
|||||  
DB 70 AAGCTGTCTCTGCTGCTAGGTTTGACAGATGACACATACACAGAAAGTTTACATCAGTACA 129  
QY 44 IleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIleLysLeuGln 63  
|||||  
DB 130 ATTGGTTTGTGATTTTAAATCAGACATATAGATTAGATGAAAAACATCAAACTGCAA 189  
QY 64 Ile----- 64  
DB 190 ATTTGGGACATCGCGGCGCAGAAAGATTTCGAACCATATCATCAGTACTACAGGGA 249  
QY 65 -----GluSerPheAsnAspValLys 71  
DB 250 GCCCAGCGCATATGTTGTGTATGACGTACAGACACAGCAAGATCTTCAACAAATGTCAG 309  
QY 72 GlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLeuValGly 91  
|||||  
DB 310 CAGTGGCTTCAGAAATAGATCGTTATGCCAGTGAACGTTAACAAGTTATTGTGTAGGC 369  
QY 92 AsnLysCysAspLeuThrThrLysLysValAlaAspTyrThrAlaLysGluPheAla 111  
|||||  
DB 370 AACAAATGTATCTCACACAAACAAAGTTGTGACTACACACAGCAAGAGATTGCA 429  
QY 112 AspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAsnValGluGln 131  
|||||  
DB 430 GACTCTGTGGAAATTCATTTTGGAAACAGCGCAAGAACGCGCAAAACGTAGAGCAG 489  
QY 132 SerPheMetThrMetAlaAlaGluIleLysLysArgMetGlyProGlyAlaThrAlaGly 151  
:::|||||  
DB 490 GCCTTCATGCAATGACAGCGGAGATCAAAAAGCGAATGGCCCTGGCGCCACAGCGGC 549  
QY 152 GlyAlaGluLysSerAsnValLysIleGlnSerThrProValLysGlnSerGlyGly 171  
|||||  
DB 550 GGTCAAGAGAGAG--AACGTCAAAATCCAGAGCATCTCAGTCAAGCAAGTGTAGTGAGGA 606  
QY 172 CysCys 173  
|||||  
DB 607 TGCTGC 612

RESULT 14  
BI684554 872 bp mRNA linear EST 18-SEP-2001  
LOCUS 603306931F1 NCL\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5343151 5',  
DEFINITION mRNA sequence.  
ACCESSION BI684554  
VERSION BI684554.1 GI:15647170  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 872)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapds-remail.nih.gov](mailto:cgapds-remail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLNL1871 row: m column: 08  
High quality sequence stop: 802.

FEATURES  
Location/Qualifiers

1..872  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5343151"  
/sex="Female, virgin"  
/tissue\_type="Infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI. Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 235 a 184 c 246 g 207 t

ORIGIN

Alignment Scores:

Pred. No.: 4.94e-88 Length: 872

Score: 807.00 Matches: 172

Percent Similarity: 83.09% Conservative: 0

Best Local Similarity: 83.09% Mismatches: 35

Query Match: 91.08% Indels: 35

Gaps: 1

US-09-820-003a-2 (1-173) x BI684554 (1-872)

QY 1 MetSerSerMetAsnProGluTyrAspTyrLeuPhelLysLeuLeuIleGlyAspSer 20  
|||||  
DB 206 ATGTCAGCAATGATATCCGAAATATGATTATTATCAAGTACTTGTGATGGGATCT 265  
QY 21 GlyValAlaLysSerCysLeuLeuLeuArgPheAlaAspAspThrTyrThrGluSerTyr 40  
|||||  
DB 266 GCGGTTGGAAAGTCT 325  
QY 41 IleSerThrIleGlyValAspPheLysIleArgThrIleGluLeuAspGlyLysThrIle 60  
|||||  
DB 326 ATCAGCACAATGTGTGTGATTTCAGATACGAACTATGAGATTAGATGGAAAAACATC 385  
QY 61 LysLeuGlnIle----- 64  
DB 386 AAGCTACAGATATGCGACACAGCGCGCAAGAAAGATTTCGAACATCACTCCAGTTAT 445  
QY 65 -----GluSerPheAs 68  
DB 446 TACAGAGAGCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 505  
QY 68 nAsnValLysGlnTrpLeuGlnGluIleAspArgTyrAlaSerGluAsnValAsnLysLe 88  
|||||  
DB 506 TAAAGTTAAACAGTGGCTGAGGAGATGATCGTTAGCCAGTGAATAATGTCAACAAGTT 565  
QY 88 uLeuValGlyAsnLysCysAspLeuThrThrLysLysValAlaAspTyrThrThrAlaLys 108  
|||||  
DB 566 GTTGGTGGGAAACAAATGTATCTGACCAAGAAAGTATGATACACAAACAGCAAA 625  
QY 108 sGluPheAlaAspSerLeuGlyIleProPheLeuGluThrSerAlaLysAsnAlaThrAs 128

MON, NOV. 18 08:42:06 2002

us-09-820-003a-2.rst

Page 11

Db	626	GGAATTCGACAGTTCCTTGCGAATTCCATTTGGTAAACAGCTACTCAACAAGCAGCAAGAA	695
QY	128	nvalglucinsrphemethermetelaiaagiuleltylsysrget-glyprogiya	148
Db	686	TGTGAACAGCTTTCATGCAGTGCAGCTGAGATTAAAAGCATGGGGCTCTGAG	745
QY	148	lanrhalaagilyalagilutys-serasnvalyisilaglinserrpr-ovallaysqln	167
Db	746	CTACAGCTGGTGGTGGCCGACAGAGCTCAATTTAAATTCAGAGACTCCAGTCAACAG	805
QY	168	SerGlYcylGlcYcs 172	
Db	806	TCAGGTGAGGCTGC 820	
RESULT 15			
LOCUS	B1739625	855 bp mRNA linear EST 20-SEP-2001	
DEFINITION	603361882F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:536876 5'		
ACCESSION	B1739625		
VERSION	B1739625.1 GI:15716638		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Enkayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	N1-MGC http://mgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1992) Contact: Robert Strausberg, Ph.D. Email: ggapds-remail.nih.gov Tissue procurement: rne capro laboratory cDNA library preparation: life technologies, inc. DNA sequencing by: mgtc genomics, inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Note: LNM11928 stop: d column: 21 High quality sequence stop: 849.		
FEATURES	Location/Qualifiers		
source	1..855 /organism="Mus musculus" /db_xref="taxon:10090" /clone_image="536876" /clone_lib="NIH_MGC_94" /tissue_type="testis" /lab_host="DH10B (phage-resistant)" /note="Osgan. eye: Vector: pCMV-SPORT6; site-1: NotI. site-2: SmaSI. Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: This is a NIH_MGC Library."		
BASE COUNT	234 a 181 c 233 g 205 t	2 others	
ORIGIN			
Alignment Scores:			
Pred. No. :	6,36e-88	Length:	855
Score:	82.00	Matches:	169
Percent Similarity:	82.21%	Conservative:	2
Best Local Similarity:	81.25%	Mismatches:	35
Query Match:	90.97%	Indels:	1
DB:	13 Gaps:		
US-09-820-003A-2 (1-173) x B1739625 (1-855)			
QY	1	MetSerSerMetLsnPrGluTYAsPTyrLeuPheLysLeuLeuLeuilegLYasPser	20
Db	185	ATGTGACACATGATGCCATATGATATTATTATTCAGGTAAGTACTTGATGGCGATCT	244
QY	21	glyvalgllylsrscysleuleuleuargphmelaaspshprtyrThrGlusertyr	40

[illegible]

Search completed: November 17, 2002, 14:30:41  
Job time : 1964 secs

